A SMALL-SCALE 3D IMAGING PLATFORM FOR ALGORITHM PERFORMANCE EVALUATION

THESIS

Steven A. James, Captain, USAF

AFIT/GAE/ENG/07-01

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

AFIT/GAE/ENG/07-01

A SMALL-SCALE 3D IMAGING PLATFORM FOR ALGORITHM PERFORMANCE
EVALUATION

THESIS

Presented to the Faculty

Department of Electrical and Computer Engineering

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

In Partial Fulfillment of the Requirements for the

Degree of Master of Science in Aeronautical Engineering

Steven A. James, BSME

Captain, USAF

June 2007

AFIT/GAE/ENG/07-01

A SMALL-SCALE 3D IMAGING PLATFORM FOR ALGORITHM PERFORMANCE
EVALUATION

Steven A. James, BSME

Captain, USAF

Approved:

_____/signed/_____   _____
Guna S. Seetharaman (Chairman)   Date

_____/signed/_____   _____
Paul R. Havig (Member)   Date

____/signed/_____   _____
Ross McNutt (Member)   Date

AFIT/GAE/ENG/07-01

Abstract

In recent years, world events have expedited the need for the design and application of rapidly deployable airborne surveillance systems in urban environments. Fast and effective use of the surveillance images requires accurate modeling of the terrain being surveyed. The process of accurately modeling buildings, landmarks, or other items of interest on the surface of the earth, within a short lead time, has proven to be a challenging task. One approach of high importance for countering this challenge and accurately reconstructing 3D objects is through the employment of airborne 3D image acquisition platforms. While developments in this arena have significantly risen, there remains a wide gap in the verification of accuracy between the acquired data and the actual ground-truth data. In addition, the time and cost of verifying the accuracy of the acquired data on airborne imaging platforms has also increased. This thesis investigation proposes to design and test a small-scale 3D imaging platform to aid in the verification of current image acquisition, registration and processing algorithms at a lower cost in a controlled lab environment. A rich data set of images will be acquired and the use of such data will be explored.

To my loving and always supportive Mother…

Acknowledgments

I would like to express my sincere appreciation to my thesis advisor, Dr. Guna Seetharaman, for his guidance and support throughout the course of this thesis effort. Your academic and personal guidance was invaluable.

I am also thankful to Julia Parakkat and the professionals at the Air Force Research Laboratory, Human Effectiveness Directorate, Biosciences and Protection Division, Biomechanics Branch for the use of their facilities to conduct my image acquisition trials.

Lastly, I would like to thank my committee, Dr. Paul Havig and Dr. Ross McNutt for reviewing my thesis and seeing the value added in my research .

Steven A. James

Table of Contents

List of Figures

List of Tables

List of Acronyms

CCD	charge coupled device
CMOS	complementary metal-oxide-semiconductor
COTS	commercial off-the-shelf
DP	dynamic program
DTED	digital terrain elevation data
FCC	federal communications commission
Fps	frames per second
GIS	geographic information service
GPS	global positioning system
GUI	graphic user interface
IED	improvised explosive device
IMU	inertial measurement unit
INS	instrument navigation system
PCB	printed circuit board
SAR	synthetic aperture radar
SVS	Small Vision System
USSTRATCOM	US Strategic Command
2D	two-dimensional
3D	three-dimensional

I. Introduction

Motivation for Research

Methods of surveillance during battlefield scenarios, intelligence gathering operations, counter-drug operation and various other surveillance applications are of increasing importance in combating terrorism and other illegal activity. Accurate modeling of buildings, landmarks or other items of interest on the surface of the earth has proven to be a challenging task for many scientists and engineers. One approach of high interest to many industries and the military for countering the challenge and accurately reconstructing 3D objects is through the employment of airborne 3D image acquisition platforms.

One such focused group which has researched, developed and tested an airborne image acquisition platform was created under a program named Project Angel Fire [1]. Project Angel Fire is a joint endeavor represented by the Air Force Institute of Technology, Los Alamos National Lab and the US Strategic Command. The program has already demonstrated many advances in image acquisition, registration and processing from an airborne platform. The basic principle of operation combines a large number of cameras mounted in a single framework with a slight offset in their respective boresights. As a whole, the array of cameras covers a wide field of view; however, separately each camera independently acquires images over a narrow field of view. When combined, the camera array lends itself to be modeled as a single wide-angle camera, particularly when the image footprint on the ground is larger than the spacing between the cameras. The surveillance aircraft flies in a circular pattern above a

designated zone and persistently observes and images a large area from a steadily changing perspective. The camera system is mounted on the right side of the aircraft and positioned pointing downward. Once sufficient images have been received, an ortho-rectified image sequence is computed by swift registration of the video sequence allowing a continual awareness of the dynamic events of the scene as shown in Figure 1.
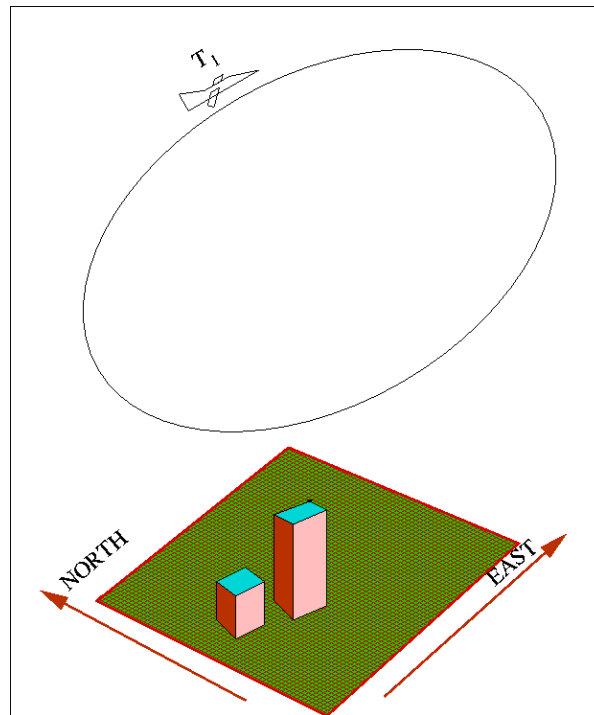
Figure 1: Project Angel Fire Concept of Operation. Airborne surveillance platform shown orbiting over a specific scene [1].

Although technology is progressing in surveillance imaging, there still remains intrinsic problems associated with image registration. A few of the problems exist with the variations in perspective, rotation and scale of the acquired surface objects as well as the high speed at which registration must be accomplished to be tactically relevant. The

underlying problems have been solved in the scientific sense; however, the massive size of the image and video frame data calls for radically new and customized algorithms to produce acceptable performance results. To a limited extent, the performance results can be sustained if 3D models of the terrain being imaged are used to steer the registration process. Therein rests a set of experimental challenges:

A) Acquiring the 3D model

B) Verifying the accuracy of the 3D model

C) Benchmarking various algorithmic tradeoffs in using the 3D model

Such comprehensive goals entail access to highly-controlled experimental evaluations involving terrain as large as several kilometers in each direction – an expensive and time consuming effort.

Another range of practical problems arise from several other conditions. One concern is the inevitable deviations in the motion of the imaging platform as a result of varying flight conditions. Weather, winds, turbulence and other atmospheric phenomena can create unfavorable platform vibrations and skewed motion which complicates the imaging solutions. Airborne platforms also have inherent errors in determining their true position relative to the earth due to errors in navigational data received from GPS or INS positioning systems. Furthermore, problems exist during the image feature extraction process including sun and sensor elevation, azimuth, shadows, occlusions, edge definition, noise and saturation of bright surfaces [3]. All of the stated issues raise scientific inquiry for the need to more accurately study these factors in a lower cost and controlled lab environment.

Research Objectives

This thesis proposes to develop and test a small-scale 3D image acquisition and test platform by which to validate a class of image registration algorithms. An essential first step is to compute the true perspective of the observed objects and estimate the instantaneous camera position and orientation with respect to a small set of known objects on the ground. This step will aid in facilitating the computation of the position and depth information in the rest of the scene and help create the digital terrain maps. The method should be robust over a wide range of perspective and scale in the encircling pattern of the overhead stereo camera platform. A small-scale lab imaging platform will also allow for image calibration, registration and processing algorithms to be tested on a ground-based truth model. Accurate 3D data of objects in the lab can easily be obtained by a simple manual measurement of the objects (X, Y and Z (depth)) and will aid in verifying imaging model algorithms being used on large-scale airborne platforms. In addition (for future work), we have incorporated a mechanism to project a stripe and facilitate direct 3D computation of all illuminated points on that stripe as recorded by the video camera. The current imaging platform was designed with the following characteristics:

A) Modular – Hardware and software components of the system should be easily constructed and allow for swift reconfiguration during operation.

B) Scalable – System operating parameters and configuration should be employable at various facilities without any major modifications.

C) Integration – Should abide by current FCC rules and regulations. Common electrical and computer outlets should be utilized.

D) Low Cost – Should use commercial off-the-shelf (COTS) hardware.

E) Easy Configuration and Maintenance – Design should allow for easy setup in a variety of settings.

The system design, operation and functional output parameters will be kept to the scope of this thesis with a look at potential uses and future upgrades.

Significance of Research

A long term goal and challenge of the Air Force and other services is persistent and pervasive surveillance. Despite a large number of research efforts and published works on image registration and object recognition, there is a critical need for a small-scale test bed which can replicate the varying conditions of airborne imaging platforms and still provide valid image sets. Due to the high complexity and range of objects in an urban environment, obtaining a verification of the perspective, location and scale of the objects or structures is a complex undertaking and, therefore, provides uncertainty in evaluating the accuracy of measurements and feature recognition. The uncertainty in predicting the true position of an object, relative to the airborne imaging platform, is not a problem unique to current Air Force projects. The same problem is evident on *Ikonos*, a commercial earth observation satellite, which was the first to collect and make public high-resolution imagery at the 1- and 4- meter resolution. Fraser [3] reports most of the published work on geometric processing of *Ikonos* imagery has surrounded the topic of insufficient accuracy in determining its full metric potential, namely the geometric accuracy of 3D positioning from stereo and multi-image coverage.

Other problems arise in the cost and approvals required to operate such a real-world platform in an urban environment. A small-scale lab imaging platform could be used as a lower expense test bed to allow for a faster verification of current algorithms used in the acquisition, registration and processing of known objects. Such a system could provide a quick turn around time in testing and developing new registration and tracking techniques.

II. Background and Theory

Overview

The purpose of this chapter is to present the background for stereo image registration, acquisition and processing in 2D and 3D scenarios. Particular attention will be focused on identifying existing approaches and deployment methods, including both past and present stereo imaging systems design. 3D target tracking systems with intelligent and automatic control systems using stereo imaging solutions are rapidly becoming more popular in government and commercial industrial applications. Stereo object tracking systems can imitate the 3D depth perception experienced in human vision by using the binocular disparity between the left and right cameras – similar to our left and right eyes. In the case of an airborne surveillance platform, as an aircraft circles above an area of interest, it acquires a steady stream of video images of varying perspective of fixed assets on the ground. Any two images separated by a relatively short time between their acquisitions will form the basis for stereo analysis, and thus a 3D perception of the observed scene.

Several low cost and economic systems will be described and a brief history of the design and development of the CCD camera and its significance in the field of 3D imaging systems will be covered. The feasibility of developing a small-scale imaging platform as a verification tool for detecting, locating and tracking an object in a framework such as Project Angel Fire, will be discussed and demonstrated.

Historical Background

A wide array of stereo imaging systems exist in various government and commercial marketplaces. Although the concepts for stereo and machine vision in manufacturing dates back to the 1930's [4], the demand for real-time imaging acquisition and processing systems didn't really begin until the mid-1960's when computer technology began displaying the speed and efficiency attractive to potential markets. In 1970, Dr. Willard Boyle and Dr. George Gomez of Bell Labs developed the world's first solid-state video camera or CCD, which is still used today in many products including digital cameras, camcorders, high-definition television, security monitoring, medical endoscopy, modern astronomy and video conferencing applications [4]. The newly discovered technology demonstrated the transmission of an electric charge along the surface of a semiconductor called the photoelectric effect. The photoelectric effect (or Hertz effect), commonly described by scientists [5], is a phenomena which takes place after exposing a metallic surface to electromagnetic radiation that is above a certain threshold frequency specific to the material and its surface condition. A current is produced when the photons are absorbed. Conservation of energy principles illustrate that as the energy of the incident photon is absorbed by the electrons it can escape from the material surface with a finite kinetic energy called photoelectricity. A CCD receives a charge from this photoelectronic energy and commonly reacts to 70% of the incident light versus 2% on a photographic type film [6]. The CCD camera then transforms these patterns of light into electrical signals. First, a capacitor array collects an image projected by a lens, allowing each capacitor to accumulate an electric charge proportional to the intensity of the light at that location. A two-dimensional array (video and still

cameras) captures the whole image or a rectangular portion of it while a one-dimensional array (line-scan cameras) captures a single slice of the image. Once the array has been exposed to the image, a control circuit causes each capacitor to shift its contents to its neighbor. The charge is converted into a voltage once the last capacitor in the array dumps its charge into an amplifier. The control circuit, after several repetitions, changes the entire contents of the array into a varying voltage, which it samples, digitizes and stores in memory [6]. An appreciation of CCD sensitivity [7] can be seen in Figure 2 showing the quantification of different sources of lux or illumination.

Table 1: Lux (Illumination) Quantitative Comparisons.

Luminance	Example
0.00005 lux	Starlight
1 lux	Moonlight
10 lux	Candle one foot away
400 lux	A brightly lit office
400 lux	Sunrise or sunset on a clear day.
1000 lux	Typical TV studio lighting
1000 lux	Level capable of producing small shifts in the human biological clock
10000 lux	Level capable of resynchronizing the human biological clock to a new schedule
32000 lux	Sunlight on an average day (min.)
100000 lux	Sunlight on an average day (max.)

The development of the CCD camera made a significant impact on stereo imaging and the science of creating the perception of a 3D image or model from separate 2D images. It is well known in this discipline that by taking two or more 2D images from various directions and transforming between the world coordinates and the image coordinates, a

3D profile of an object can be created. Several optical systems have used CCD technology to advance the field of stereo imaging and applications as shown in the following vision system descriptions.

Vision Systems

System 1: 3D Vision Sensor with Multiple CCD Cameras [8]

A high speed, accurate 3D visual inspection system was developed for printed circuit boards (PCBs) without using expensive or sophisticated optical equipment. Using up to 17 CCD cameras arranged in a hemispheric pattern, various optimal combinations were used to detect the precise 3D positions of components on a PCB after applying stereo image matching algorithms. Stereo image matching was resolved using the brightness distribution between a two camera combination with the use of a two step DP method beginning at the pixel level followed by an 8 times sub-pixel expansion. The desired accuracy (1 mm) and rapid processing time (< 10 ms) for PCB board inspection was achieved and lends to the technology of rapid 3D image acquisition at a low cost without the use of expensive, high-tech equipment.

System 2: Adaptive 3D Target Tracking and Surveillance Scheme based on Pan/Tilt-Embedded Stereo Camera System [9]

Stereo vision has also aided in the development of an adaptive real-time intelligent face tracking system. In this system, sequential stereo image pairs were acquired at a rate of 30 frames per second (fps), at a resolution of 320 x 240 pixels, allowing for a geometric measurement of distance and the 3D coordinates. By

incorporating a robotic pan/tilt system the developers were able to create an algorithm centered on the subject of interest and record position displacement data that was in turn relayed to the pan/tilt system for tracking. Standard deviation of the position displacement of the target in the horizontal and vertical directions were low at an average of 1.5 pixels, while the error ratio between the measured and computed 3D coordinate values of the target was 0.5% on average [9]. This significant research implies real-time target tracking using an active vision stereo imaging system is attainable and adds value to investigating the feasibility of creating a small-scale test bed to validate various other sensor data.

Relevant Research

Project Angel Fire [1]

Project Angel Fire is a USSTRATCOM requested and sponsored airborne surveillance platform being developed and tested to counter the IED and urban warfare issues. In collaboration with Los Alamos National Lab and AFIT, the program aims to provide real-time tactical situational awareness of city-size urban environments. USSTRATCOM requests that the surveillance platform be able to identify suspicious targets and track them in time and space with the ability to communicate the information to operational users in rapid succession. In addition, the platform needs to have the ability to characterize IED events during the pre- and post- detonation phases. All detected events must be able to be played forward and backward in time for higher level analysis. Figure 2 [1] shows the Angel Fire conceptual approach to target, acquire and relay tactical information. In short, Project Angel Fire desires to deploy an airborne

platform to a medium-size urban environment to loiter for extended periods of time and relay images in high resolution. Of particular interest to this thesis is the feasibility of Project Angel Fire to acquire and register the images. The development of a small-scale imaging test bed, which essentially emulates the image acquisition process of an Angel Fire airborne platform, could prove to be a viable time and cost saver in verifying the accuracy and overall effectiveness of current image processing algorithms.
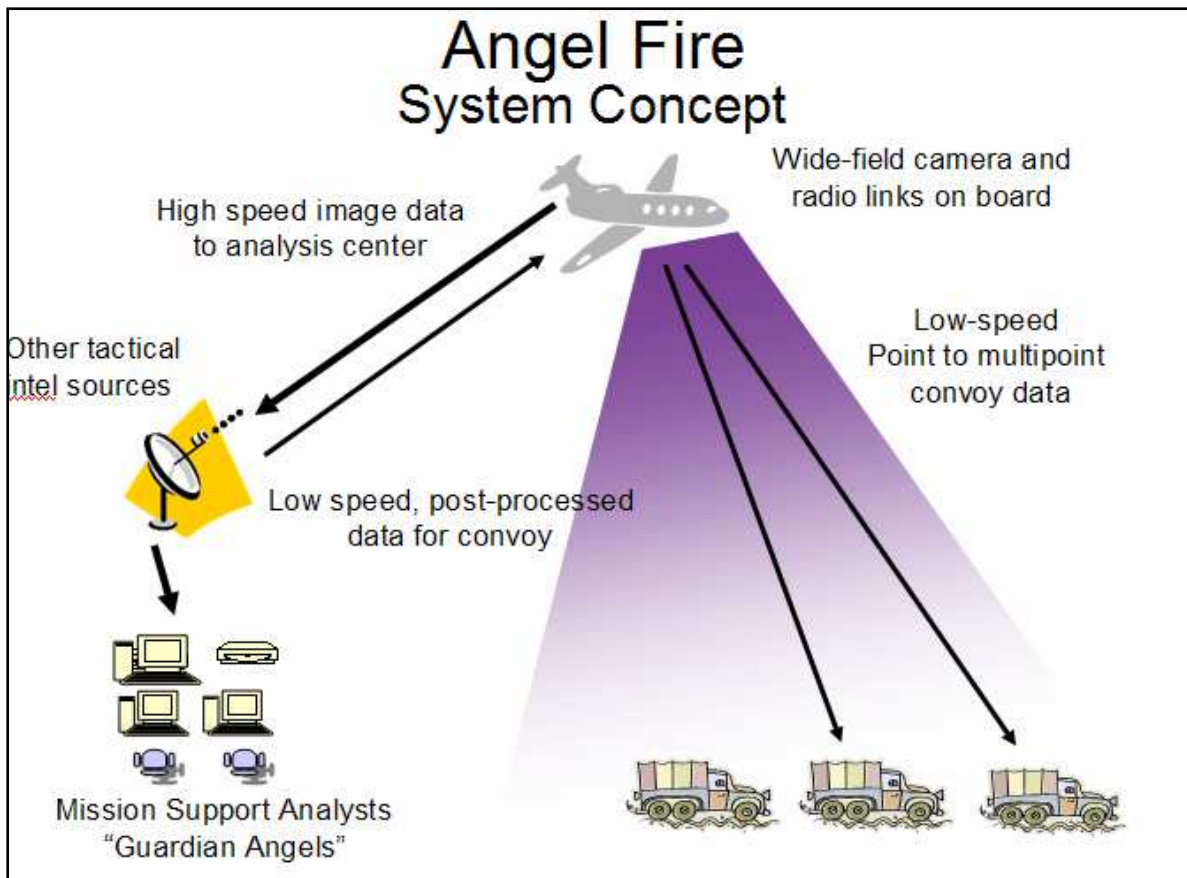
Figure 2: Project Angel Fire. Airborne surveillance platform and associated components for image and data relay [1].

Chapter Summary

Details of an extensive literature search provided a historical and current view of research efforts and a sample of the applications in stereo imaging relevant to this thesis. The background and operation of the CCD camera was described and several examples of its uses were shown with the center of interest on Project Angel Fire, a current and relevant Air Force project. A number of universities, including Stanford, are also focusing on similar problems under the broad topics of persistent surveillance, video-SAR and light-field imaging. The discussion illustrated that stereo imaging is not a new concept; however, its uses and implementation into various new areas of science and technology could provide innovative solutions to many imaging problems.

III. Methodology

Overview

This chapter will discuss the materials and methods by which the proposed benchmark imaging research was conducted. First, a brief description of the research facility and the equipment used will be covered. Next, a description of the small-scale stereo imaging platform setup and its associated hardware is given. To finalize the chapter, an explanation of the test setup and procedures is detailed and followed by a methodology conclusion.

Human Effectiveness Facility

The research was performed at the Air Force Research Laboratory, Human Effectiveness Directorate, Biosciences and Protection Division, Biomechanics Branch at Wright Patterson Air Force Base, Ohio, in Building 824. The facility has a spacious area on the ground level used for various experiments and was an ideal place to set up the imaging platform and network of computers. Also located in this area of the building was a heavy duty 2000 pound max load capacity winch which was used to raise and lower the stereo imaging platform (approx 50 lbs) for data collection. The maximum height of the cameras at the operating limit of the winch in this particular facility was 6.5 ft, high enough to capture images of the objects placed in the view of the camera pair through a 360 degree rotation. Other facilities may offer different winch options for variations in the image acquisition heights.

Setup Parameters

Two CCD cameras captured the field objects in monochrome stereo and stored the information in two groups (left and right cameras), via an IEEE 1394 interface, into
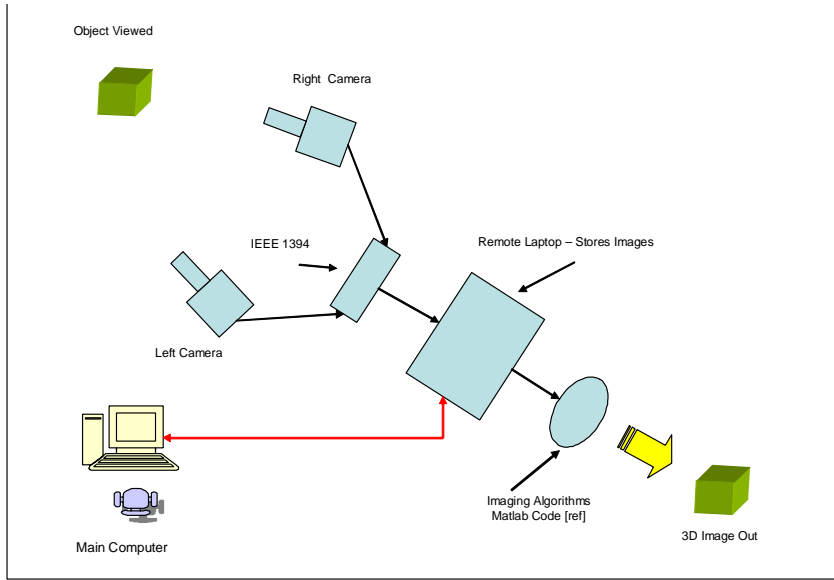
Figure 3: Imaging Platform Flowchart. Relay of 2D image data through electronic components from the input object to the output display.

the memory of a remote laptop computer. The setup and flow of operations is described in Figure 3. The remote laptop computer on the imaging platform was wirelessly operated from a main computer at 54 Mbps to download and process the image information received.

The first set of images captured was of a test field for calibration purposes and the second set of images captured was of a "mock scene" described later. A full 360 degree rotation of the cameras took place for each set of images, in essence to simulate one overhead circle of an airborne platform loitering above an urban environment. The

images were taken under ambient room lighting conditions and the left and right images for each set were acquired in real time. The image size and baseline were also varied between the two sets of images captured to allow for a more diverse image set for analysis. Table 2 outlines the parameters used in each of the two different baseline image sets.

Table 2: CCD Camera Parameters. 10 ft and 8 ft baseline camera characteristics.

Parameter	10 ft Baseline	8 ft Baseline
Left Camera Height (mm)	1993.5	1993.5
Right Camera Height (mm)	1962.15	1962.15
Exact Baseline (mm)	2898.775	2305.05
Captured Image Pixel Size	640 x 480	320 x 240
Calibration Images Captured (single 360 deg rotation)	20	21
Mock Scene Images Captured (single 360 deg rotation)	1600	800

Imaging Platform

The design of the platform was created with several considerations in mind as outlined in the introduction. First, the platform needed to be easily constructed using market competitive or off-the-shelf components and have the ability to be transportable to facilitate future research in stereo imaging. Second, the platform needed to be robust enough to withstand being disassembled and reassembled or have components which could be easily replaced quickly at a low cost. Finally and most importantly, the platform needed to be designed to capture images in stereo combination through a 360 degree

rotation.   Several iterations of the design have been explored and a final design was selected which best met the above stated criteria and is shown in Figure 4. In general, the
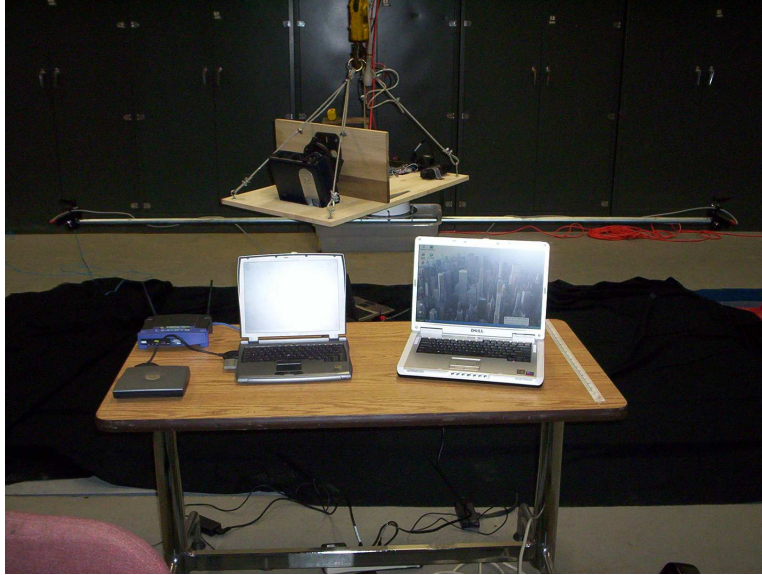
Figure 4: Small-scale Imaging Platform. Completed design in background with associated computer operating network shown in front.

design consists of a base structure, modified ceiling fan, adjustable camera baseline rod, two CCD cameras, laptop tub and a digital projector (for future work).

Base Structure

The base structure and mounting surface of the platform consists of a 2 x 3 x ¾ inch section of plywood as shown from both sides in Figure 5 and Figure 6. A more detailed description of their orientation on the platform will be described in each component's subsection of this thesis. The platform is held from each corner by plastic wrapped steel cable attached to hooks mounted through the base board of the platform.
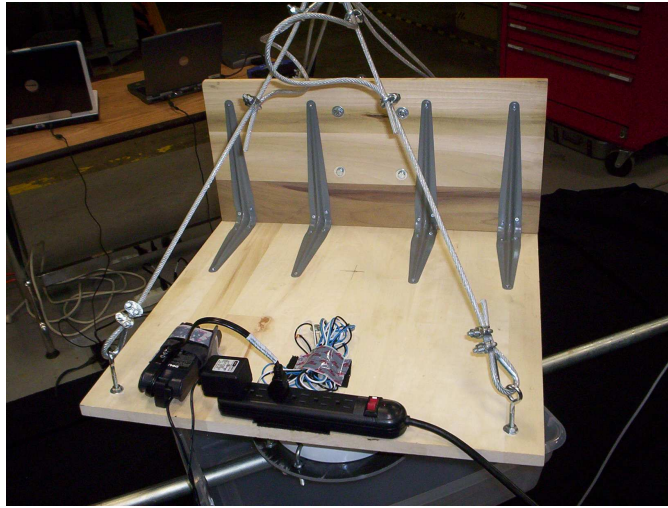
Figure 5: Imaging Platform Base (top left view). Image platform shown with steel cable supports and associated electrical connectors for the remote laptop computer and CCD cameras.
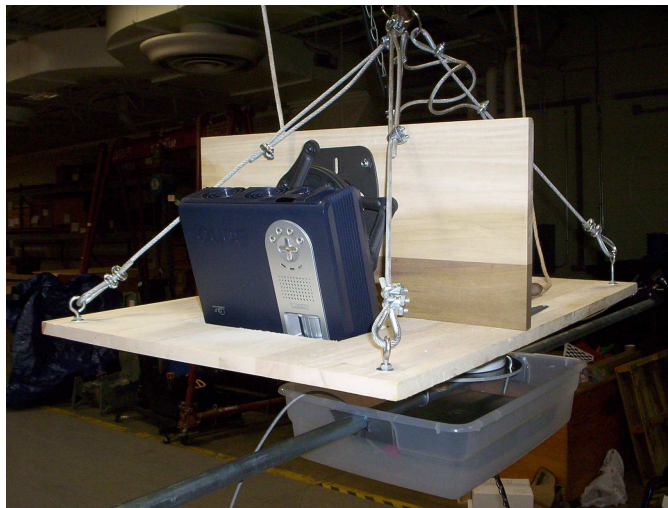

Figure 6: Imaging Platform Base (top right view). Image platform shown with steel cable supports, digital projector and remote laptop computer tub.

The heavy duty cables and mounts ensured the platform did not become a safety hazard during the raising or lowering throughout the image acquisition process. Two steel rings

are also attached to each pair of cables (at opposite ends of the base board) and will allow for either a central mounting point at the approximate center-of-gravity or for 2 separate mounting points depending on the facility used.

Modified Ceiling Fan

The modified ceiling fan (Figure 7) and the 10 ft adjustable camera baseline rod were designed to allow for a smooth circular rotation of the 2 CCD cameras and provided the best COTS alternative for the ease of assembly and low cost. The ceiling fan readily
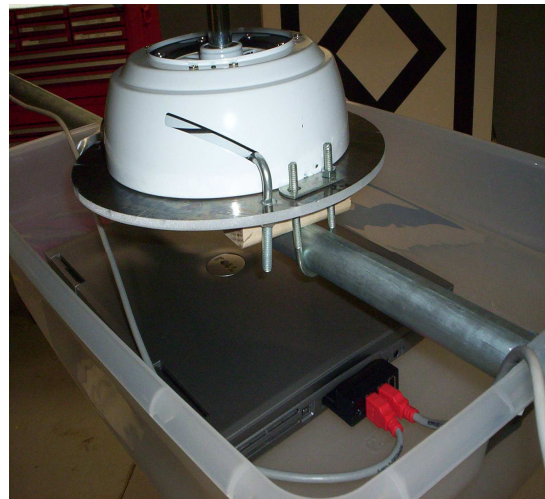
Figure 7: Modified Ceiling Fan. Left image shows the fan attachment to the bottom of the base platform. Right image shows the circular base plate added to the fan with U-clamps to hold the 10 ft. camera rod.

consists of the internal mechanisms, such as pre-sealed ball bearings and a rotating shaft, which would sustain a long life of repeated use. The ceiling fan has also been left with its electrical components intact to allow for future modifications or studies where power may be applied for rotation.

Camera Baseline Rod and CCD Cameras

The adjustable camera baseline rod is a simple 10 ft steel hollow tube. Several types of cameras and mounting devices can be used at any point along the rod allowing for easier baseline adjustments and more flexibility in the image acquisition process. Figure 8 shows the Videre Systems STH-MDCS-VAR CCD cameras [10] used throughout the experimentation and their orientation along the camera baseline rod.

Figure 8: CCD Cameras and Camera Baseline Rod. CCD cameras and their relative size (left). CCD camera mounted on the baseline rod and attached to the IEEE 1394 fire wire.

The CCD cameras are low-power, compact digital stereo heads with an IEEE 1394 (fire wire) interface. Each camera consists of two 1.3 megapixel progressive scan CMOS imagers with their own fire wire peripheral interface module. The CMOS imagers are capable of up to a 1280 x 1024 pixel image in a monochrome ½ inch format. The imagers are fully controllable through the fire wire interface and the user can set and adjust several camera characteristics including exposure, gain and decimation.

The dynamic range, sensitivity, and noise characteristics of the CMOS imagers allow for a wide-range of image acquisition. Each camera is equipped with standard CS-mounted lenses for use with interchangeable optics and each are electronically synchronized to one another, as well as to an 8 KHz clock on the IEEE 1394 interface, allowing images to be captured at exactly the same time. The stereo cameras can be accessed and operated on MS Windows 98SE/ME/2000/XP and for Linux 2.4.x kernels and utilize software written by SRI International [11]. Camera calibration, stereo correlation and their results can also be accessed and manipulated through the use of the software package.

Remote Laptop Computer Tub

A standard 5 gallon plastic storage container (Figure 9) was modified to hold a laptop functioning as the interface between the CCD cameras. A 1 inch hole was cut out of each end of the tub allowing the camera baseline rod to pass completely through. The tub and rod were then mounted to the ceiling fan using standard hardware as shown in Figure 10.

Digital Projector

A BenQ PB6200 Digital Projector was also mounted to the imaging platform as seen in Figure 11. . The projector can act as a stipe-gird projector to aid in the selection of edge points for image registration. The projector was added to provide for future research into 3D image acquisition. A rectangular portion of the base platform plywood was removed to allow for variations in the projection orientation with respect to the scene below. Lim [12] conjectures that by projecting parallel light planes onto a scene they
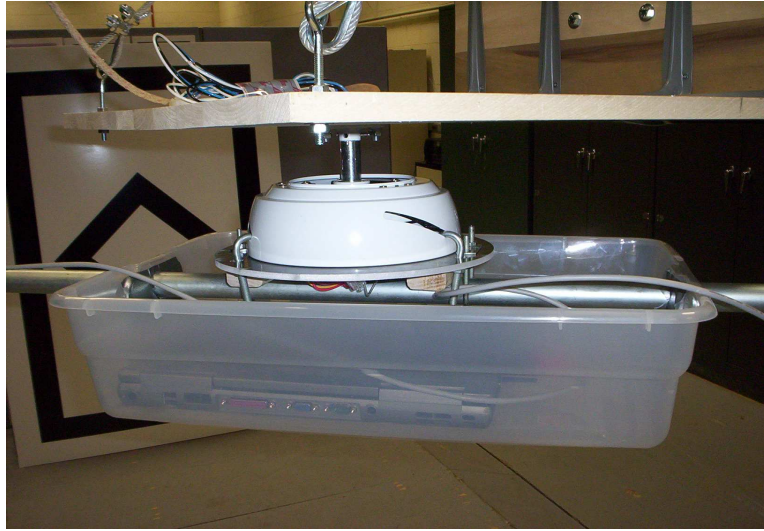
Figure 9: Remote Laptop Computer Tub. Remote laptop tub and IEEE 1394 fire wire camera interface. The modified ceiling fan is also shown attached to the base platform.
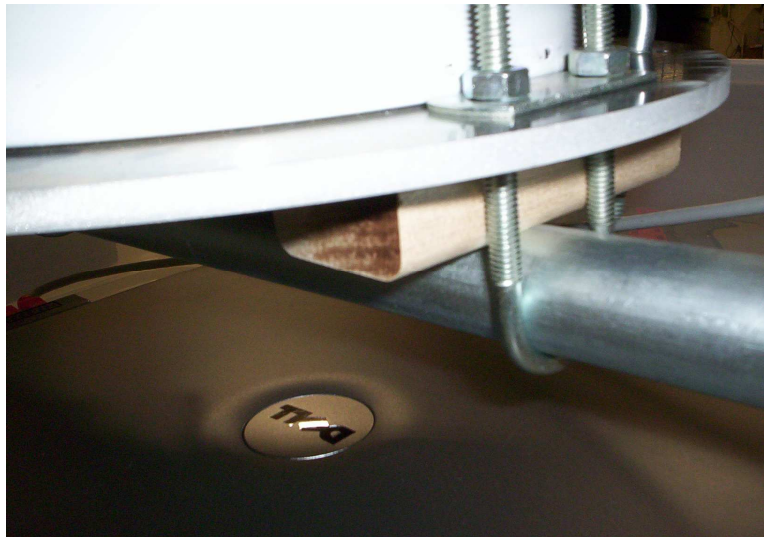

Figure 10: Baseline Camera Rod Mounts. U-clamps with spacer for baseline camera rod. The rod holds the laptop tub, laptop and IEEE 1394 fire wire camera interface.

will appear as a set of broken straight lines in the viewed images. Discontinuities along these straight lines correspond to normal discontinuities on the underlying surfaces and the edge points can then be more easily extracted. The mounting bracket for the projector was attached in such a way to allow for rotation of the projector and better align its field of projection to the scene below. All normal projector functions are available for use and operation.

Figure 11: Digital Projector. BenQ PB6200 projector mounted to the vertical support. Cutout shown in the base platform allows for adjustments to the projector field of transmission.

Calibration

Acquiring 3D images via a standard stereoscopic system proceeds through three basic procedures: calibration, registration and processing. During calibration, the normal process of obtaining 3D images from 2D information begins by aligning two or more

images of a scene. Several different methods have historically been used in calibrating a stereo camera system [13]. Usually one image will be the reference image and the other image will be matched pixel by pixel to the corresponding points in the reference image. By identifying the position of a known object in the reference image, the identities of the remaining objects and their position and orientation in another image can be determined. The cameras must be calibrated before the images can be matched in a stereo combination. Reconstruction of the 3D structure in an image requires solving equations connecting the coordinates of a point in 3D space to the coordinates of the corresponding point in the image. The goal of camera calibration is to recreate a perfect pinhole camera with exactly the same parallel optical axes and focal length. In reality, most cameras are imperfect due to lens distortion, uneven focal lengths and misaligned optical axes. Camera calibration determines the intrinsic and extrinsic parameters of the stereo system which are used in compensating for their imperfections. The intrinsic parameters correct for lens distortion and uneven focal length while the extrinsic parameters determine the spatial offset of the two cameras, the stereo baseline and any deviation from the parallel optical axis. In other words, the intrinsic parameters are the parameters necessary to link the pixel coordinates of an image point with the corresponding coordinates in the camera reference frame and the extrinsic parameters are the parameters that define the position and orientation of the camera reference frame with respect to a known world reference frame [14]. The intrinsic and extrinsic parameters can then be used to adjust the camera images into a standard position as seen by two pinhole cameras with parallel optical axes. The calibration approach described in the next section is well known in stereo imaging

practices.   Table 3 defines the intrinsic and extrinsic parameters and the associated

variables that were used.   Figure 12 shows an example of the physical relationship

between the world reference frame and the camera reference frame.

Table 3: Intrinsic and extrinsic calibration variables and their definitions.

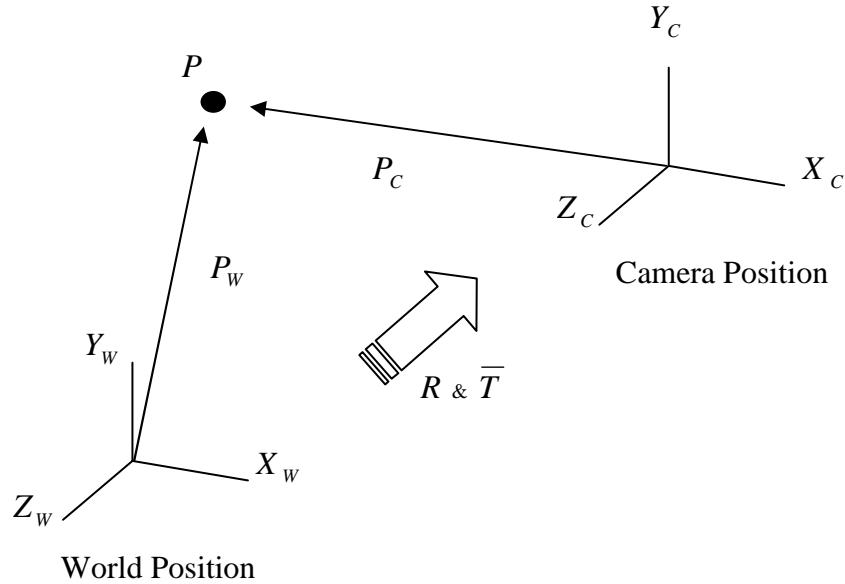| Camera Calibration Parameters			
Intrinsic Parameters	Definition	Extrinsic Parameters	Definition
f	Focal length	R	3 x 3 Rotation matrix
S_x	Horizontal pixel size	$\vec{T}$	3-D Translation vector
S_y	Vertical pixel size		
O_x	X-coord of image center		
O_y	Y-coord of image center		
k	Radial distortion coefficient		

Figure 12: Camera to World Coordinate Transformation. Point P in relation to the Camera (X_c, Y_c, Z_c) and World (X_w, Y_w, Z_w) coordinate frames.

25

The calibration method chosen involves measuring the image coordinates

$$(u_i, v_i), i = 1, 2, \cdots, N$$

Where u_i = x-coordinate of the image plane

v_i = y-coordinate of the image plane

of several well known 3D points:

$$(X_i, Y_{,i} Z_i), i = 1, 2, \cdots, N$$

Then, we seek to solve a linear system of homogenous equations in 12 mutually constrained unknowns: $q_1, q_2, q_3, \cdots, q_{12}$. One standard approach to solving homogeneous equations is to set one of the unknowns as unity and then solve the system of equations for one less variable, followed by a suitable rescaling process. These unknowns are referred to in the P matrix below such that:

$$P = \begin{bmatrix} q_1 & q_2 & q_3 & q_4 \\ q_5 & q_6 & q_7 & q_8 \\ q_9 & q_{10} & q_{11} & q_{12} \\ 0 & 0 & 0 & \varepsilon \end{bmatrix} \equiv \begin{bmatrix} q_{1x} & q_{1y} & q_{1z} & T_x \\ q_{2x} & q_{2y} & q_{2z} & T_y \\ q_{3x} & q_{3y} & q_{3z} & T_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{1}$$

We can then form a set of linear equations:

$$\begin{bmatrix} X_i & Y_i & Z_i & 1 & 0 & 0 & 0 & 0 & -u_i X_i & -u_i Y_i & -u_i Z_i \\ 0 & 0 & 0 & 0 & X_i & Y_i & Z_i & 1 & -v_i X_i & -v_i Y_i & -v_i Z_i \end{bmatrix} \begin{bmatrix} q_1 \\ q_2 \\ \vdots \\ q_{11} \end{bmatrix} = q_{12} \begin{bmatrix} u_i \\ v_i \end{bmatrix}. \tag{2}$$

As previously stated, we set one of the variables equal to unity (q_{12}). The other variables can then be solved and allows us to exploit the constraints to estimate the scale factor.

In this case, the scale factor is designated as "ε" and we have an equation of the form:

$$\mathbf{Aq=b}$$

Where

$$\mathbf{q}=\left(\mathbf{A^t A}\right)^{-1}\mathbf{A^t b}.$$

The scale factor, ε, can be determined from:

$$\varepsilon^2 = q_9^2 + q_{10}^2 + q_{11}^2$$

Using this value of ε we compute:

$$q_{1x} = \frac{q_1}{\varepsilon}, q_{1y} = \frac{q_2}{\varepsilon}, \cdots, T_y = \frac{q_8}{\varepsilon}, \quad \text{and} \quad T_z = \frac{1}{\varepsilon}$$

$\mathbf{A}$ is a $2N \times 11$ matrix and $\mathbf{c}$ is a $2N \times 1$ vector. The intrinsic and extrinsic parameters can now be extracted from this matrix $\mathbf{Q}$. Further insight into the derivation [14] reveals that:

$$\begin{bmatrix} q_{1x} & q_{1y} & q_{1z} & T_x \\ q_{2x} & q_{2y} & q_{2z} & T_y \\ q_{3x} & q_{3y} & q_{3z} & T_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \equiv \begin{bmatrix} \alpha_x \mathbf{r}_1^t + u_0 \mathbf{r}_3^t & \vdots & \alpha_x t_x + u_0 t_z \\ \alpha_y \mathbf{r}_2^t + v_0 \mathbf{r}_3^t & \vdots & \alpha_y t_y + v_0 t_z \\ \mathbf{r}_3^t & \vdots & t_z \\ \mathbf{0}^t & \vdots & 1 \end{bmatrix}$$

Where, $\quad \alpha_x = \dfrac{f}{\Delta u}$, and $\alpha_y = \dfrac{f}{\Delta v}$

defines the relationship between the focal length and pixel dimensions. A common practice is to choose either the pixel dimension or the focal length as a ground-truth among the other ground-truths (namely the world coordinates of the control points). The

27

terms: (u_0, v_0) represent the true optical center expressed in the image coordinates (digitized grid). The vector

$$\mathbf{t} = (t_x, t_y, t_z)^t$$

is the position of the camera in the true world coordinate system. Finally, $\mathbf{r}_1, \mathbf{r}_2$ and $\mathbf{r}_3$ represent the direction cosines of the x, y and z axes of the camera respectively. These values can be extracted as follows:

$$\mathbf{r}_3 := \mathbf{q}_3 \quad \text{and} \quad t_z := T_z.$$

$$\alpha_x := \|\mathbf{r}_3 \times \mathbf{q}_1\|, \quad \text{and} \quad \alpha_y := \|\mathbf{r}_3 \times \mathbf{q}_2\|$$

$$u_0 := \mathbf{r}_3 \bullet \mathbf{q}_1; \quad \text{and,} \quad v_0 := \mathbf{r}_3 \bullet \mathbf{q}_2$$

$$t_x := \frac{(T_x - u_0 t_z)}{\alpha_x}; \quad \text{and} \quad t_y := \frac{(T_y - v_0 t_z)}{\alpha_y}.$$

It is important to note how the derivations were made.

First let:

$$\begin{bmatrix} X_w \\ Y_w \\ Z_w \end{bmatrix} = \begin{bmatrix} r_{1X} & r_{2X} & r_{3X} \\ r_{1Y} & r_{2Y} & r_{3Y} \\ r_{1Z} & r_{2Z} & r_{3Z} \end{bmatrix} \begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} + \begin{bmatrix} T_X \\ T_Y \\ T_Z \end{bmatrix}$$

where the columns $\mathbf{r}_1, \mathbf{r}_2$ and $\mathbf{r}_3$ of the matrix $\mathbf{R}_c$ represent the direction cosines of the X, Y, and Z axes of the camera coordinate system and T is the position of the camera measured from the world coordinate system. Typically, the matrix $\mathbf{R}_c$ and vector $\mathbf{T}_c$ are known through information from the IMU and GPS respectively. This equation is useful in computing the coordinates of targets from the images but with additional constraints.

28

Its dual form, however, is more useful for camera calibration. The dual form is written

as:

$$
\begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} = \begin{bmatrix} r_{1X} & r_{1Y} & r_{1Z} \\ r_{2X} & r_{2Y} & r_{2Z} \\ r_{3X} & r_{3Y} & r_{3Z} \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix}
$$

where the vector **t** represents the location of the world-coordinate frame origin, measured

with respect to the camera coordinate system. Thus:

$$
\mathbf{t} = -\mathbf{R}_c{}^{\mathbf{t}}\mathbf{T}_c .
$$

The previous equation can be expressed as a single (invertible) linear transformation of

the form:

$$
\begin{bmatrix} x_c \\ y_c \\ z_c \\ 1 \end{bmatrix} = \begin{bmatrix} r_{1X} & r_{1Y} & r_{1Z} & t_x \\ r_{2X} & r_{2Y} & r_{2Z} & t_y \\ r_{3X} & r_{3Y} & r_{3Z} & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} .
$$

The perspective projection of the overall system lets us conclude that:

$$
\frac{x_i}{z_i} = \frac{r_{1X}.X_i + r_{1Y}Y_i + r_{1Z}Z_i + t_x}{r_{3X}.X_i + r_{3Y}Y_i + r_{3Z}Z_i + t_z}, \text{ and } \frac{y_i}{z_i} = \frac{r_{2X}.X_i + r_{2Y}Y_i + r_{2Z}Z_i + t_y}{r_{3X}.X_i + r_{3Y}Y_i + r_{3Z}Z_i + t_z}.
$$

These two equations can be solved numerically with at least 6 corresponding image point

pairs.

First, we introduce a normalized retinal plane (Figure 13), called the $uv-$plane such that:

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} \equiv \begin{bmatrix} u \\ v \\ f \end{bmatrix} \text{ suggesting, } \frac{u}{x} = \frac{v}{y} = \frac{f}{z} = \frac{1}{s}$$

where *"s"* is an unknown scale factor corresponding to the exact distance of the object from the camera.
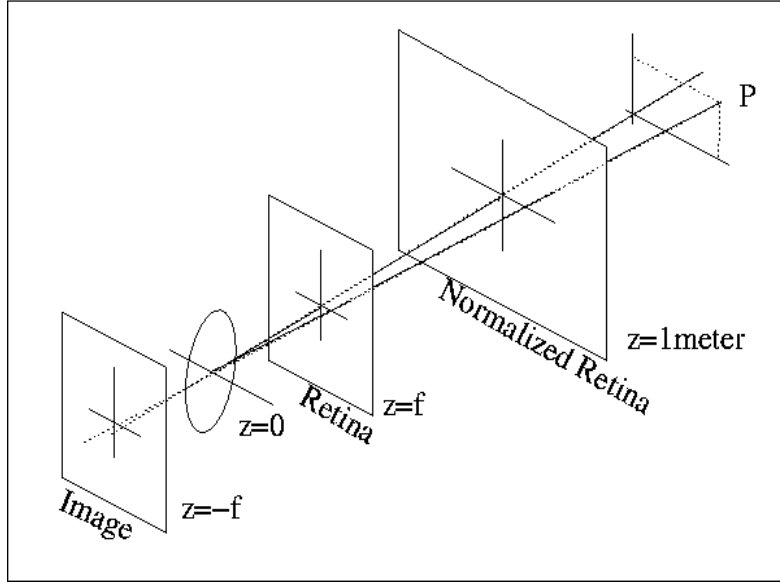

Figure 13: Measured pixel coordinates in the image plane.

Note that all values of $s > 0$, since the depth information is lost and the retinal plane is in front of the lens at (z = f), whereas the exact CCD-plane is at (z = -f).

Then,

$$\frac{u}{xf} = \frac{v}{yf} = \frac{1}{z} \qquad \text{and,} \qquad \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

The plane $[u,v,1]^T$ instantiates that there is another plane parallel to the image plane and the retinal plane, however this time with $z = 1$. Let the image grid be on this plane. Now, we define:

$$u = (m_u - u_0)\Delta_u$$

and

$$v = (m_v - v_0)\Delta_v$$

where (u_0, v_0) is the optical center on the $z = 1$ plane and (u_0, v_0) is the location of the same optical center on the image grid measured in pixels and is subsequently dimensionless (Figure 14).
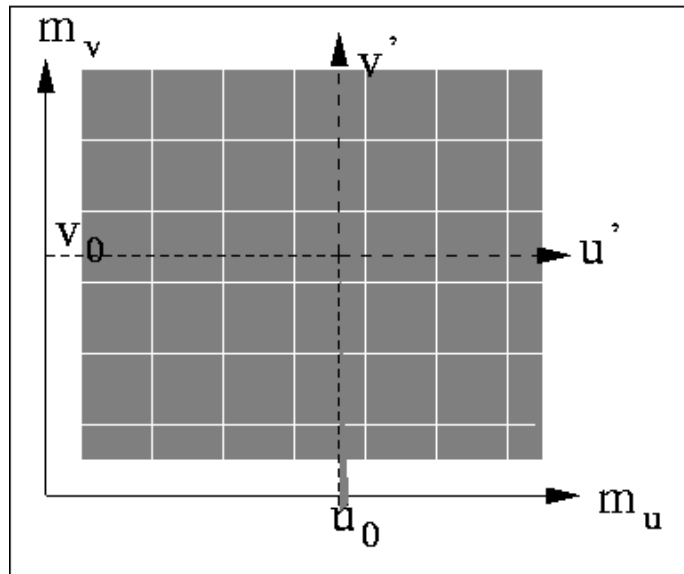

Figure 14: Planes involved in deriving the calibration model.

If we substitute $k_u = \Delta_u^{-1}$, and $k_v = \Delta_v^{-1}$, then,

$$
\begin{bmatrix} m_u \\ m_v \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & u_0 \\ 0 & 1 & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} k_u & 0 & 0 \\ 0 & k_v & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} k_u & 0 & u_0 \\ 0 & k_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix}
$$

Where,

$$
\begin{cases}
\Delta_u = \dfrac{1}{k_u} \quad \text{pixel width} \\[4mm]
\Delta_v = \dfrac{1}{k_v} \quad \text{pixel height} \\[4mm]
\tau_u = \Delta_u u_o \quad \text{x-coord of optical center} \\[2mm]
\tau_v = \Delta_v v_o \quad \text{y-coord of optical center}
\end{cases}
.
$$

Then,

$$
\begin{bmatrix} m_u \\ m_v \\ 1 \end{bmatrix} = \begin{bmatrix} k_u & 0 & u_0 \\ 0 & k_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}
$$

Thus,

$$
\begin{bmatrix} m_u \\ m_v \\ 1 \end{bmatrix} = \begin{bmatrix} k_u & 0 & u_0 \\ 0 & k_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{1X} & r_{1Y} & r_{1Z} & t_x \\ r_{2X} & r_{2Y} & r_{2Z} & t_y \\ r_{3X} & r_{3Y} & r_{3Z} & t_z \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix}
$$

32

Given a point $\left(X_x, Y_w, Z_w\right)$ and its observed location (m_u, m_v) on the image plane (Figure 14), we could then write:

$$\frac{m_u}{\mathbf{q}_1{}^t\mathbf{X} + q_{14}} = \frac{m_v}{\mathbf{q}_2{}^t\mathbf{X} + q_{24}} = \frac{1}{\mathbf{q}_3{}^t\mathbf{X} + q_{34}} = \lambda, \text{ for some } \lambda > 0$$

The above description shows the manner in which equation (1) is derived. The optical center (u_0, v_0) is measured in pixel coordinates. Thus, (u_0, v_0) is a dimensionless pair of numbers indicating its position in the grid. Figure 15 shows the inertial frame of a vehicle and the associated world-based measurements. Typically, you only need the heading and pitch; however, in reality you also need roll so the analytical process continues.
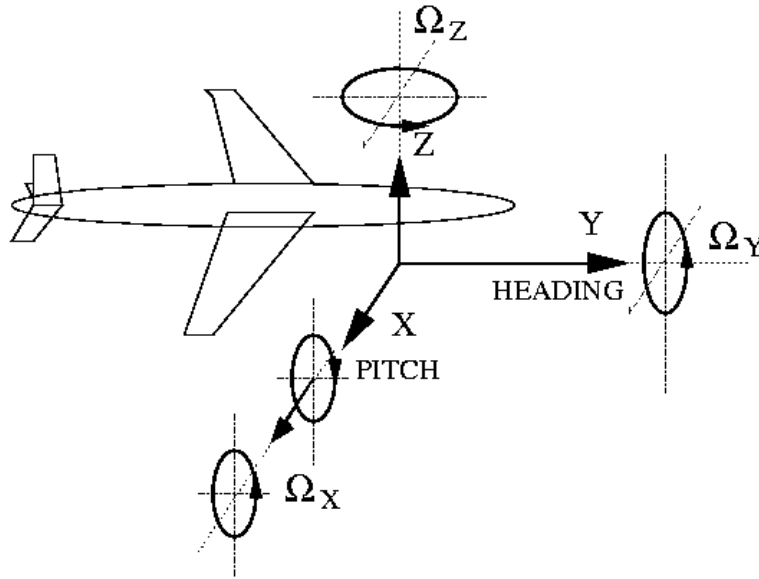

Figure 15: Inertial frame of an aircraft and the associated world coordinates.

For example in Figure 16, the derivation is based only on heading and pitch.
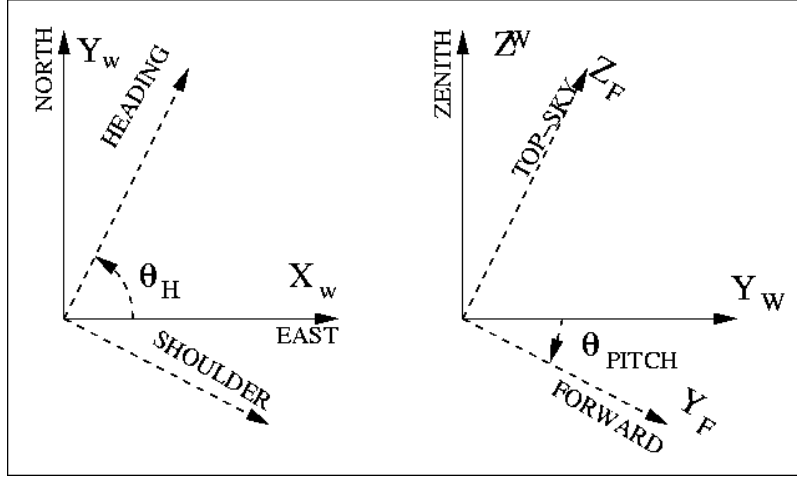
Figure 16: A pair of primary relationships between frames.

Step 1: Compute the earth-fixed coordinates of several well-known points on the area to be surveyed. This would require choosing an arbitrary origin (could be a land mark point) and at least five other points. Let these be:

$$\left(X_i, Y_i, Z_i \right), i = 1, 2, \cdots, N.$$

Step 2: Using some interactive procedure, including the possible use of an image processing toolbox (in our case the Camera Calibration Toolbox for Matlab [15]), we next locate the image coordinates of these control points in the image. Let these be:

$$\left(u_i, v_i \right), i = 1, 2, 3, \cdots, N.$$

Step 3: Form a $2N \times 11$ matrix $\mathbf{A}$ and a $2N \times 1$ vector $\mathbf{c}$ such that:

$$\mathbf{A} = \begin{bmatrix} X_1 & Y_1 & Z_1 & 1 & 0 & 0 & 0 & 0 & -u_1 X_1 & -u_1 Y_1 & -u_1 Z_1 \\ 0 & 0 & 0 & 0 & X_1 & Y_1 & Z_1 & 1 & -v_1 X_1 & -v_1 Y_1 & -v_1 Z_1 \\ X_2 & Y_2 & Z_2 & 1 & 0 & 0 & 0 & 0 & -u_2 X_2 & -u_2 Y_2 & -u_2 Z_2 \\ 0 & 0 & 0 & 0 & X_2 & Y_2 & Z_2 & 1 & -v_2 X_2 & -v_2 Y_2 & -v_2 Z_2 \\ \vdots & \vdots & \vdots & \vdots & & & & & & & \\ X_N & Y_N & Z_N & 1 & 0 & 0 & 0 & 0 & -u_N X_N & -u_N Y_N & -u_N Z_N \\ 0 & 0 & 0 & 0 & X_N & Y_N & Z_N & 1 & -v_N X_N & -v_N Y_N & -v_N Z_N \end{bmatrix}, \text{ and } \mathbf{c} = \begin{bmatrix} u_1 \\ v_1 \\ u_2 \\ v_2 \\ \vdots \\ u_N \\ v_N \end{bmatrix}.$$

Step 4: Compute:

$$\tilde{\mathbf{q}} = \left(\mathbf{A^t A} \right)^{-1} \mathbf{A^t c};$$

Or

$$\tilde{\mathbf{q}} = \left(\mathbf{A^t \Lambda A} \right)^{-1} \mathbf{A^t \Lambda c};$$

where, Λ defines the confidence of each observation by a non-zero weight.

Step 5: Compute: $\mathbf{q}$ from $\tilde{\mathbf{q}}$ using the scalar ε such that,

$$q_9^2 + q_{10}^2 + q_{11}^2 = 1; \quad \text{and} \quad \varepsilon q_{12} = 1.$$

Step 6: Compute and verify if $q_1^2 + q_2^2 + q_3^2 = 1$ and if $q_5^2 + q_6^2 + q_7^2 = 1$. If this holds true, then we can safely conclude that the image pixel dimensions are equal to unity and the optical center is exactly at the grid center of the image. However, this is seldom the case and we move on to Step 7.

Step 7: Compute:

- $\mathbf{r}_3 := \mathbf{q}_3$ and $t_z := T_z$.

- $\alpha_x := \|\mathbf{r}_3 \times \mathbf{q}_1\|$, and $\alpha_y := \|\mathbf{r}_3 \times \mathbf{q}_2\|$. Note: $\alpha_u = \dfrac{f}{\Delta_x}$; and, $\alpha_v = \dfrac{f}{\Delta_y}$.

- $u_0 := \mathbf{r}_3 \bullet \mathbf{q}_1$ and $v_0 := \mathbf{r}_3 \bullet \mathbf{q}_2$ where $(\tau_u = u_o / \Delta_u, \tau_v = v_o / \Delta_v) = (\tau_u, \tau_v)$ are the locations of the optical center of the camera.

- Construct the matrix: $\mathbf{R} = [\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3]$ from the 3x1 vectors $\mathbf{r}_1, \mathbf{r}_2$ and $\mathbf{r}_3$.

- Compute $\mathbf{T_c} = -\mathbf{R}_c \mathbf{t_c}$

Step 8: Repeat Step 7 for each camera.

Step 9: At this point, we distinguish between the vehicle frame coordinate system, the world earth-fixed coordinate system and the camera coordinate system. The $\mathbf{R}$ matrix computed in Step 6 is a product of two matrices, -- $\mathbf{R}_{F|w} \cdot \mathbf{R}_{c|F} = \mathbf{R}_{c|w}$ in which the former matrix is known through the IMU, and the latter matrix is intrinsic to how the camera has been fitted on the vehicle frame. Thus, compute:

$$\cdot \mathbf{R}_{c|F} = \mathbf{R}_{F|w}^{-1} \mathbf{R}_{c|w}$$

and

$$\mathbf{T}_{c|F} = \mathbf{R}_{F|w}^{-1} (\mathbf{T}_{c|w} - \mathbf{T}_{F|w})$$

where $\mathbf{T}_{F|w}$ is the onboard GPS reading – indicating the position of the vehicle frame origin with respect to the IMU. The values $\mathbf{R}_{c|F}$ and $\mathbf{T}_{c|F}$ are intrinsic to each camera. They depend on the relative orientation and position of each camera to the vehicle frame. In general, the GPS and IMU positioning solutions should be kept closer together. If not, the homogeneous transformations are likely to be prone to anisotropic errors in

36

displacements and locations of targets with respect to the platform. Also, note that the optical center (u_o, v_o) and its equivalent image-grid-location (τ_u, τ_v) are intrinsic to the camera once a lens has been fitted and are most sensitive to changes when using an auto-focus and/or an auto-aperture system. Radial distortions have not been considered and would involve a more elaborate interpretation of **q.**

Videre Camera Calibration

The camera calibration of the Videre stereo system utilized a typical stereo pair of CCD cameras setup for capturing and processing video images. A video capture board or frame grabber then digitized the video streams into the main memory of the remote laptop computer located in the laptop tub (Figure 9). This experimental setup used the Small Vision System (SVS) program from SRI International [11] as the graphic user interface (GUI) during the image capture process. Then, using the Camera Calibration Toolbox for Matlab functions [15], stereo pairs were created between the left and right cameras and used as input arguments into the Matlab code for the camera calibration. Once calibration was complete, the input arguments can be used to further process the images as defined by a particular user. The method chosen for this calibration analysis, however, utilizes a unique setup. A common procedure for camera calibration involves viewing a planar calibration target from several different orientations while a pair of stereo cameras remains stationary. Conversely, in this calibration, the stereo pair will be rotating and capturing images as it is moves through 360 degrees while suspended above a large checkerboard pattern as shown in Figure 17.
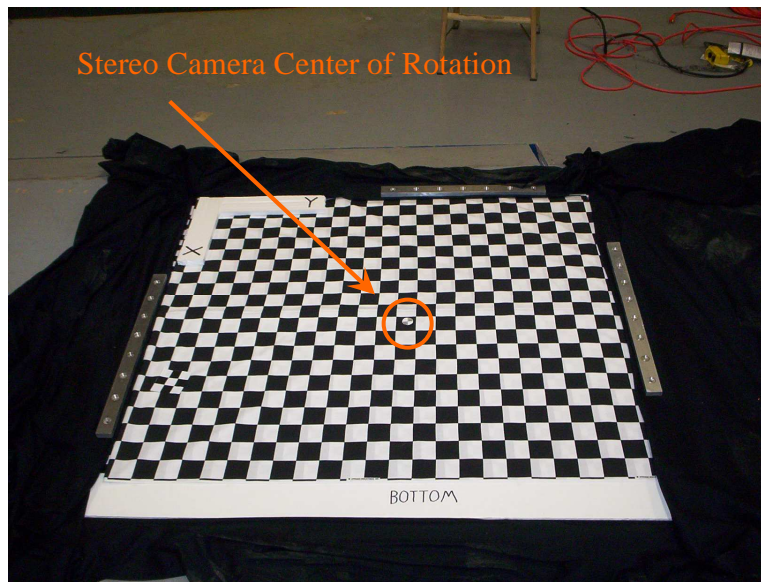
Figure 17: Calibration Checkerboard. Top left is the X & Y origin Camera center of rotation is shown.

The checkerboard overall dimensions are approximately 4 x 4.5 feet. The exact overall dimensions are irrelevant to the camera calibration; however, the exact pixel dimensions (in mm) of each checkerboard square are very important in determining the intrinsic and extrinsic parameters of the stereo pair. Figure 18 shows the dimensions of each square to be 2.125 inches or 53.95 mm.
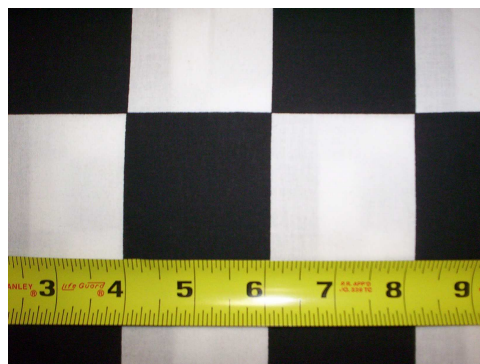
Figure 18: Calibration Checkerboard Dimensions. Squares are 53.975 mm x 53.975 mm on each side.

Other important characteristics of the calibration setup were previously listed in Table 2. The following analysis represents the calibration procedure used with the stereo cameras for both the 10 ft baseline and the 8 ft baseline, although only the 10 ft baseline calibration process will be discussed. A complete detailed list of the calibration steps can be found in the Camera Calibration Toolbox for Matlab program [15]. First, the images were separated into 2 groups: left camera calibration images and right camera calibration images. Each set of left and right images were calibrated separately and were then combined for a stereo pair calibration. Next the images were loaded into the memory of a PC by defining a base name and image format (bitmap in our case). Once loaded, a complete set of left and right calibration images are produced as shown in Figure 19 and 20.
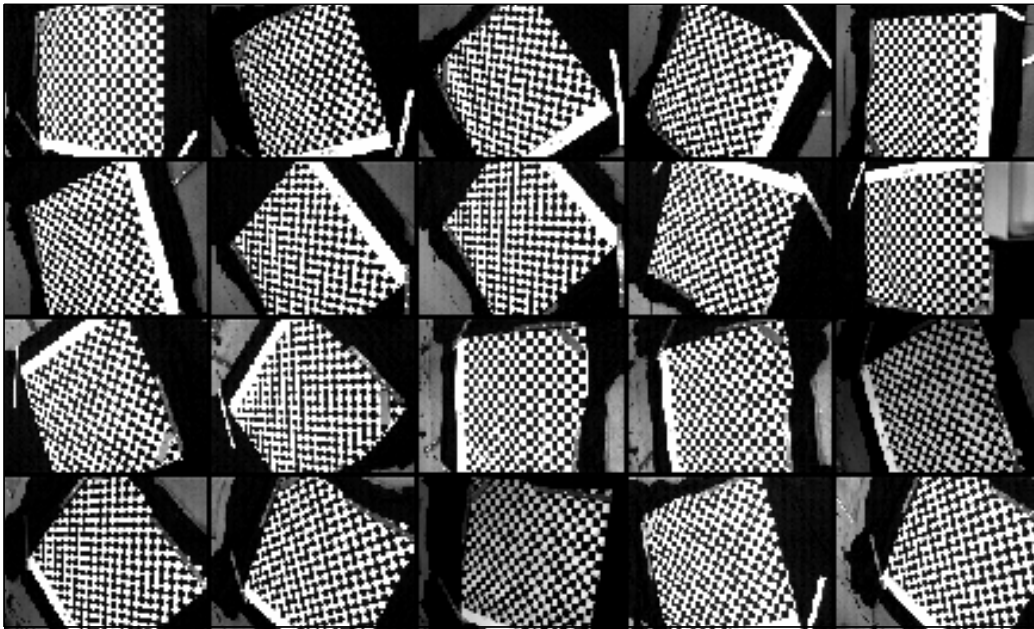

Figure 19: Left Calibration Images. 10 ft baseline calibration images. Cameras rotated through a 360 degree circle.
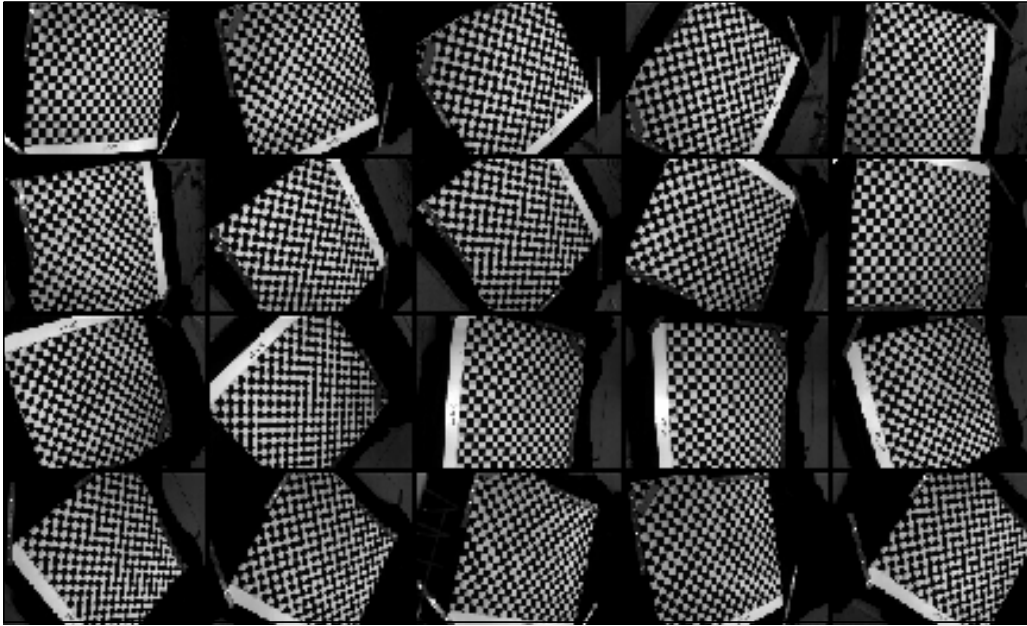
Figure 20: Right Calibration Images. 10 ft baseline calibration images. Cameras rotated through a 360 degree circle.

Next, the overall grid corners were selected for each of the left and right images. As seen in Figure 19 and 20, not all checkerboard squares are visible in each image. Therefore, a calibration pattern had to be selected that would be visible in all calibration images. A window search size of 11 x 11 pixels was used to manually select four corner points from each image to define the largest commonly viewable checkerboard pattern. The selected corner points are shown in Figure 21. The large green "O" in each image's upper left corner represents the selected origin. The green X and Y axes are also displayed. After the outermost corner points were defined, an automatic counting mechanism (or manual selection if desired) will count the number of squares within the defined parameters once the specific square size is defined.   In this case, each square has a size of 53.95 mm x 53.95 mm.

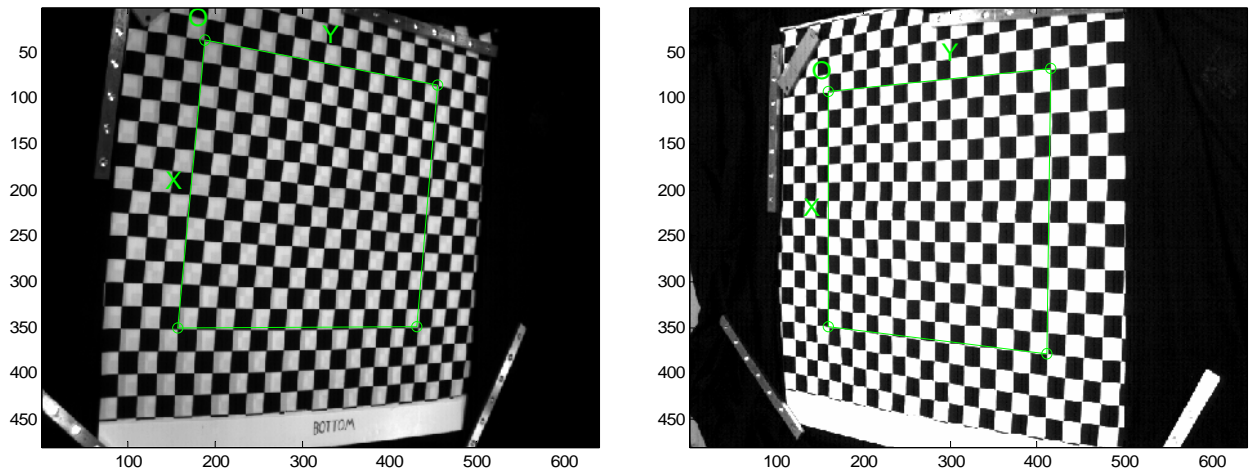**Figure 21: Calibration Corner Points (left and right cameras). 10 ft baseline manually selected corner points. X & Y axes and origin (all in green) are shown on the checkerboard. Pixel dimensions are shown on the outside X & Y axes (640 x 480).**
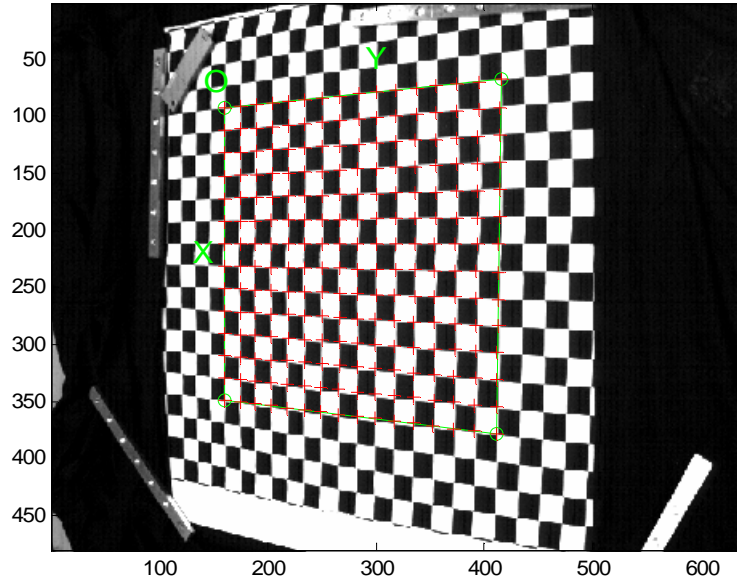

Figure 22: Prediction of entire checkerboard corner points (left camera). 10 ft baseline computer generated corner points. Red crosses should be close to corner points.
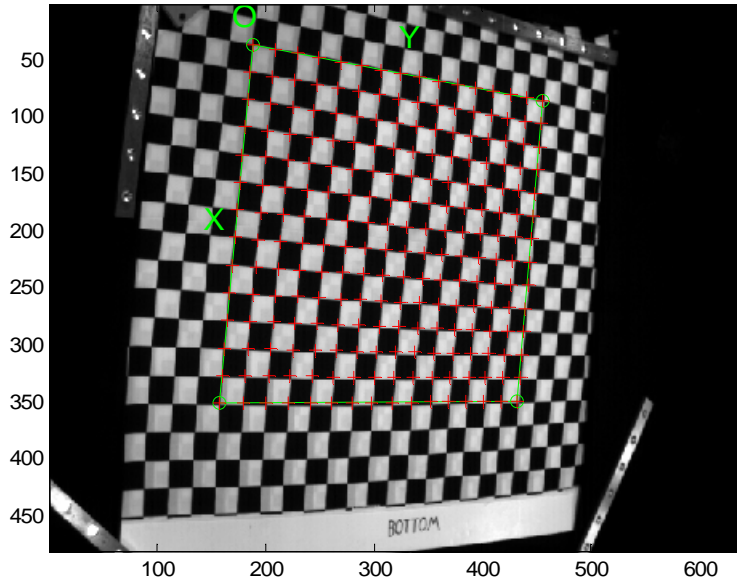
**Figure 23: Prediction of entire checkerboard corner points
(Right camera). 10 ft baseline computer generated corner
points. Red crosses should be close to corner points.**

The program will then predict where each of the image corners are for each square within

the user defined pattern as shown in the left image and right images in Figures 22 and 23,

respectively. The option now exists to accept the program generated corner points (if

they are close to the actual image corners) or enter a distortion factor to account for the

radial distortion of the images. In this case, the corner points selected in Figure 22 and

23 are close to the actual image corners and the program generated each corner point to

an accuracy of about 0.1 pixels [15] for each image as shown in Figures 24 and 25. After

the corner extraction was completed for each image, the intrinsic and extrinsic parameters

were calculated and the results are shown in the Results and Conclusions section of this
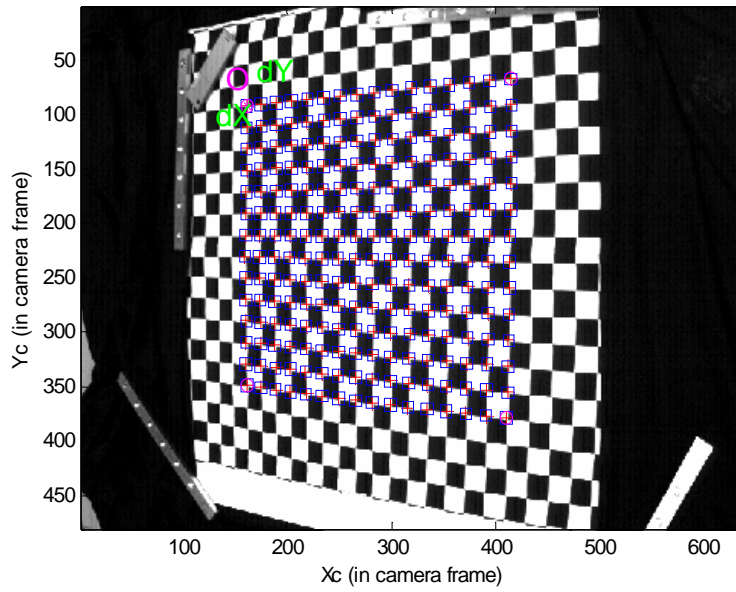
thesis (Chapter 4).

Figure 24: Extracted corner points (left camera). 10 ft baseline computer generated corner points. Corner points are accurate to approximately 0.1 pixels [15].
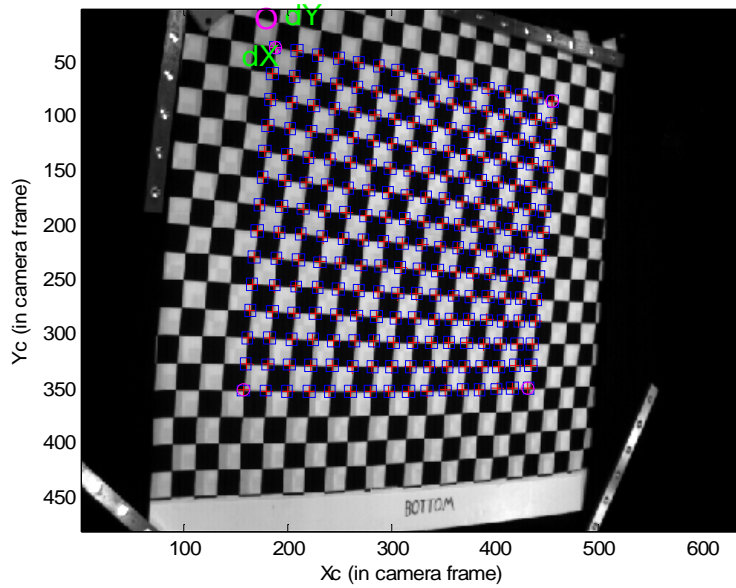

Figure 25: Extracted corner points (Right camera). 10 ft baseline computer generated corner points. Corner points are accurate to approximately 0.1 pixels [15].

Registration

       Once the cameras were calibrated a mock scene was created to simulate an urban environment and capture a robust set of images which could also be used as an analysis tool for verifying the accuracy of imaging algorithms in future work. Figure 26 shows the objects to be used in the scene and Table 4 shows each object's dimensions and orientation in relation to the origin of the X & Y coordinate system visible in the top left corner of each image. Objects of different sizes, shapes and orientations were selected for the imaging process which were in high contrast with the black background. The data in Table 4 is the "ground-truth" data (described in the Research Objectives) for verification of the depth information in future work during the image registration process. Figure 27 shows the setup of the objects within the mock scene. A healthy set of 1600 images (2 x 400 each right and left cameras) was taken at the 10 ft baseline and another 1600 images (2 x 400 each right and left cameras) at the 8 ft baseline. The 10 ft baseline images were taken with a 640 x 480 resolution and the 8ft baseline images were taken with a 320 x 240 resolution to allow for a more diverse image set to analyze. Figure 28 shows a captured left and right pair of mock scene images at the 10 ft baseline and Figure 29 shows a pair of images captured at the 8 ft baseline. By knowing the coordinates of each object with respect to the origin and each object's dimensions (ground-truth data), a relationship can be made as to the accuracy of the spatial dimensions (2D dimensions plus depth) extracted from the image registration and processing of the mock scene.
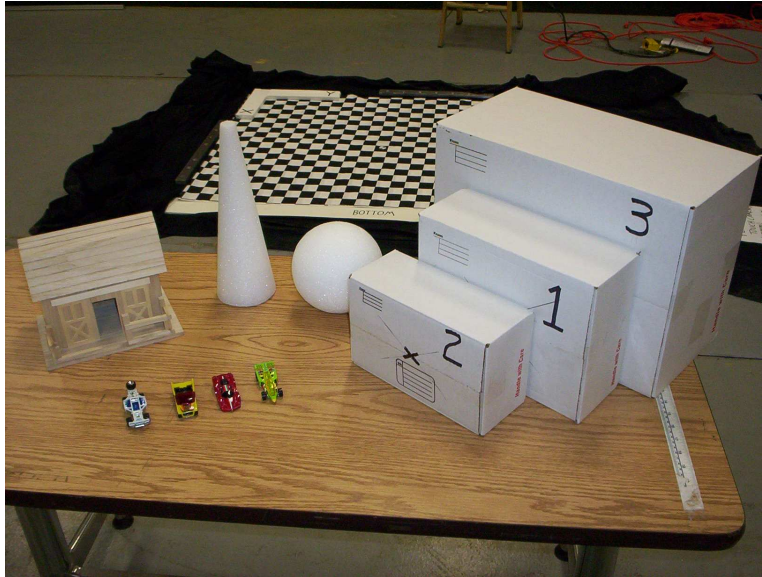
Figure 26: Mock Scene Objects. Objects of various sizes and shapes with known dimensions. Objects will serve as ground-truth data points.

Table 4: Mock Scene Object Parameters. Data will serve as ground-truth information for verification of the algorithm accuracy from the acquired imaging information.

	Object Parameters (mm)			Orientation to Origin	
Object	Length	Width	Depth	X (mm)	Y (mm)
Box #3	406.400	304.800	203.200	5483.860	21653.183
Green Car	76.200	25.400	19.050	2822.575	10725.785
Box #2	241.300	152.400	101.600	8225.790	12096.750
Yellow Car	63.500	31.750	19.050	11209.655	5161.280
White Car	82.550	31.750	25.400	14435.455	19677.380
Sphere	139.700	0.000	0.000	15645.130	12379.008
Cone	304.800	101.600	0.000	17580.610	4435.475
Box #1	292.100	222.250	107.950	20241.895	19032.220
Birdhouse	209.550	152.400	177.800	22822.535	10161.270
Red Car	69.850	31.750	12.700	22419.310	6612.890

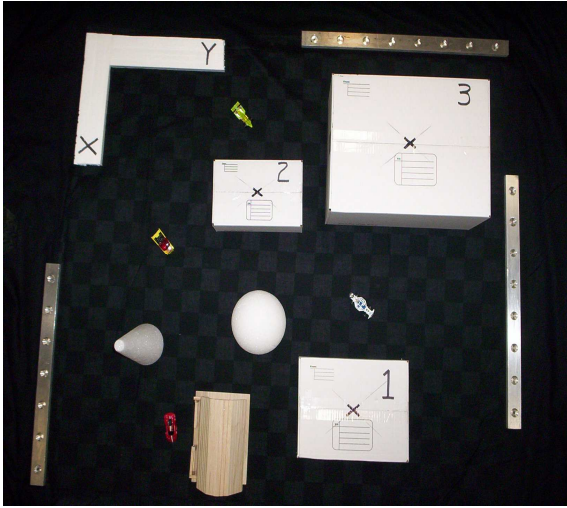**Figure 27: Mock Scene Objects. Objects shown with various orientations to the X and Y origin (upper left corner) in high contrast with the black background. Overhead view (left image) and a 3D perspective (right image).**
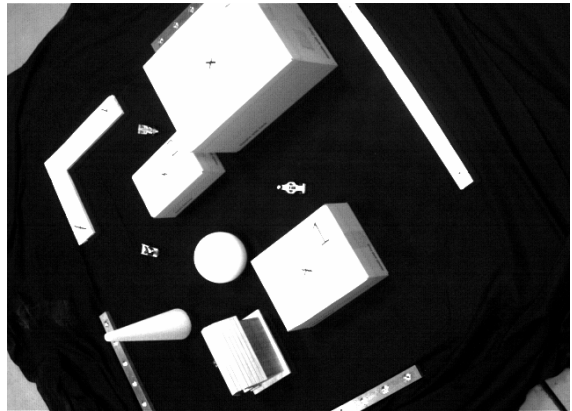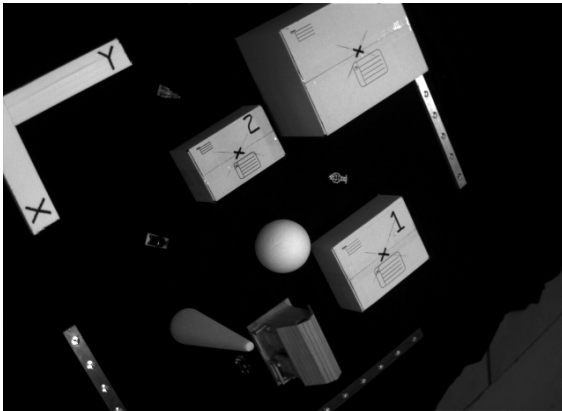

Figure 28: Mock Scene Images (left and right cameras). 10 ft baseline at 640 x 480 resolution. Images taken at 3.75 fps through a 360 degree rotation. 800 pairs of images captured.
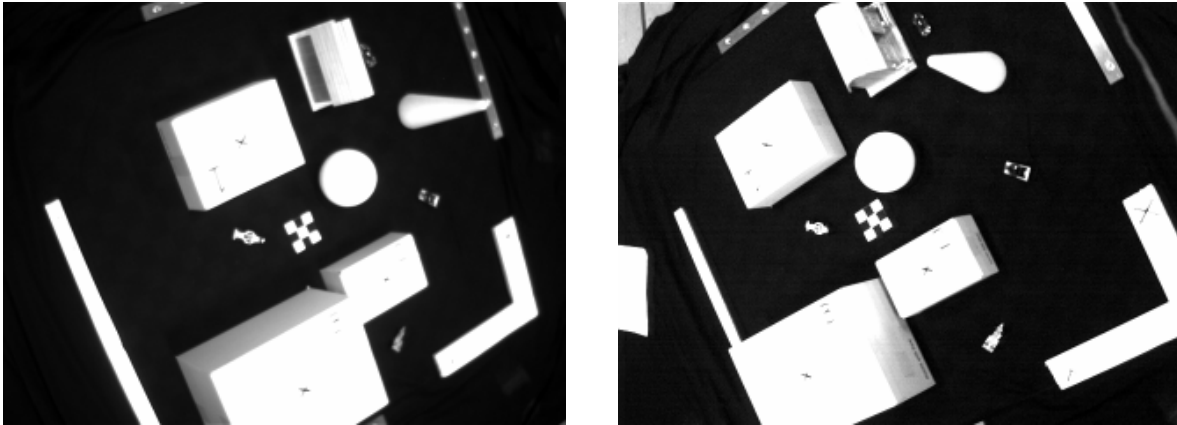
Figure 29: Mock Scene Images (left and right cameras). 8 ft baseline at 320 x 240 resolution. Images taken at 3.75 fps through a 360 degree rotation. 400 pairs of images captured.

Chapter Summary

A brief description of the facility used with the small-scale imaging platform is defined. The components of the platform are also characterized in greater detail and several images were provided which show the individual component characteristics and the overall design at completion. Next, the theory of the imaging platform operation is outlined, demonstrated and discussed. An explanation of the calibration mathematics, process and the associated parameters were presented and displayed. An overview of the image registration process and the applicability and importance in verifying the ground-truth data acquired from real-world platforms is shown.

Results and Discussion

Overview

An explanation of the Matlab code [15] used and the results are shown for the calibration of both the 8 ft and 10 ft baseline. An interpretation of image registration is given as well as a more narrow focus on the type of image registration required for the validation of data from airborne imaging platforms. The need for a small-scale imaging platform for valuable data collection and analysis will be demonstrated.

Calibration

The two steps used in the calibration process with the Camera Calibration Toolbox for Matlab [15] are initialization and nonlinear optimization. Excluding lens distortion, the initialization process computes a closed form solution for the calibration parameters, while the nonlinear optimization minimizes the total reprojection error over all of the calibration parameters. The calibration parameters used are described in the Camera Calibration Toolbox for Matlab [15]. The 10 ft baseline calibration process converged to within 3/1000 of a pixel (2D) within 5 iterations and the 8 ft baseline calibration converged to within 3/1000 of a pixel (2D) in 4 iterations. The results of the calibration parameters of each baseline are shown on pages 49 and 50.

10 ft Baseline – Left Camera

Calibration results after optimization (with uncertainties):

Focal Length: fc = [1823.50410   1394.40639] ± [97.47758   29.97083]

Principal point: cc = [319.50000   239.50000] ± [0.00000   0.00000]

Skew: $alpha\_c$ = [0.00000] ± [0.00000]   => angle of pixel axes =

90.00000 ± 0.00000 degrees

Distortion: kc = [-1.03345   5.36270   -0.01571   -0.00210 0.00000]

± [0.10726   3.43103   0.00141   0.00229 0.00000]

Pixel error: err = [0.39710   0.41444]

10 ft Baseline – Right Camera

Calibration results after optimization (with uncertainties):

Focal Length: fc = [1730.65149   1221.23700] ± [151.98471   49.31769]

Principal point: cc = [656.19070   261.39494] ± [0.00000   0.00000]

Skew: $alpha\_c$ = [0.00000] ± [0.00000]   => angle of pixel axes =

90.00000 ± 0.00000 degrees

Distortion: kc = [-0.47011   -0.77875   -0.02006   -0.09979 0.00000]

± [0.04773   0.38888   0.00157   0.00589 0.00000]

Pixel error: err = [0.42156   0.38355]

8 ft Baseline – Left Camera

Calibration results after optimization (with uncertainties):

Focal Length: $fc = [\ 1195.24826\quad 794.18894\ ] \pm [\ 140.06847\quad 33.04137\ ]$

Principal point: $cc = [\ 159.50000\quad 119.50000\ ] \pm [\ 0.00000\quad 0.00000\ ]$

Skew: $alpha\_c = [\ 0.00000\ ] \pm [\ 0.00000\ ]$ => angle of pixel axes =

$$90.00000 \pm 0.00000 \text{ degrees}$$

Distortion: $kc = [\ -1.23510\quad 17.84935\quad -0.02496\quad 0.01252\quad 0.00000\ ]$

$$\pm [\ 0.23437\quad 14.02673\quad 0.00209\quad 0.00317\quad 0.00000\ ]$$

Pixel error: $err = [\ 0.22021\quad 0.19861\ ]$

8 ft Baseline – Right Camera

Calibration results after optimization (with uncertainties):

Focal Length: $fc = [\ 1150.04832\quad 857.92774\ ] \pm [\ 83.82160\quad 23.21282\ ]$

Principal point: $cc = [\ 159.50000\quad 119.50000\ ] \pm [\ 0.00000\quad 0.00000\ ]$

Skew: $alpha\_c = [\ 0.00000\ ] \pm [\ 0.00000\ ]$ => angle of pixel axes =

$$90.00000 \pm 0.00000 \text{ degrees}$$

Distortion: $kc = [\ -1.20554\quad 10.54276\quad -0.02945\quad 0.03526\quad 0.00000\ ]$

$$\pm [\ 0.21396\quad 12.15550\quad 0.00282\quad 0.00433\quad 0.00000\ ]$$

Pixel error: $err = [\ 0.20286\quad 0.19415\ ]$

Registration

Image registration is the important task of transforming different sets of data, taken at different times, from different viewpoints, by different sensors into one coordinate system and is a crucial step in all post-imaging analysis techniques. The past few decades have flourished with many new developments in image registration and the growth of image acquisition devices. In just the last ten years, the Institute of Scientific Information reports that over 1000 papers have been published in the topic of image registration [16]. Most methods of image registration [17] are commonly separated into two main registration classes:

  a) Feature-based

  b) Area-based

The two main registration classes are described by Zitova [16] as follows:

Feature-based methods first focus on detection of objects within the image that are easily discernable and detectable in both images. Major surface or terrain objects make excellent features for extraction (forests, lakes, coastlines, rivers etc). Once the features are detected the next step in the registration process is to match the various common points between the separate images.

Area-based methods of image registration are more concerned with the feature-*matching* step, rather than first detecting certain details as in the feature-*based* method. Without detecting the specific features in an image, the area-based method uses "window" type segments of an image, or even an entire image, to match areas, regions or illumination and intensities which are similar in the images.

In general, 2D registration refers to relating two different stereo images in some manner which correlates both images to the same coordinate system, while 3D recovery refers to extracting the 3D information from video images which are essentially 2D. The 'recovery' portion is of the extraction of the depth information which was lost in the image processing of 2D images. 2D to 3D registration refers to taking a 2D image (with the loss of depth) and matching it against a known 3D scene and extracting the 3D information of objects in the scene, which may not have been in the original scene model. A good example of 2D to 3D registration would be of surveillance images from a downtown area where the model usually includes buildings and terrain information without the pedestrians, vehicles and other dynamic objects. The chief task in video surveillance includes:

1) 2D registration over time.

2) Forming incremental 3D recovery solutions from 2D registration and stereo analysis.

3) 3D registration of the imprecise, incremental and partial 3D data over time so as to build a useful 3D model of the scene.

4) Using one or more 2D images as they become available and partially mapping each against the 3D model to help understand and analyze the 3D dynamics of the underlying 3D scene.

This thesis specifically defines image registration as the mapping of the same points between two or more 2D images and relating that information to a known 3D scene (the mock scene setup). The mock scene setup has the "ground-truth" data built in since all of the objects have known dimensions, scale, rotation and position. Similarly, the image registration used in Angel Fire relates the distinct features of 2D images to those features of a known 3D or reference scene (typically DTED or GIS data) as shown in Figure 30. The challenge in image registration remains to overcome the loss of depth information inevitably found in optical imaging systems. The design of a low cost imaging platform with which to more rigorously study these challenges is essential and can quickly provide a variety of image sets to analyze. Therefore, the proposed small-scale imaging platform could provide valuable insight and allow for a better analysis of the accuracy of the image data required by Project Angel Fire or other airborne platforms with similar imaging profiles.
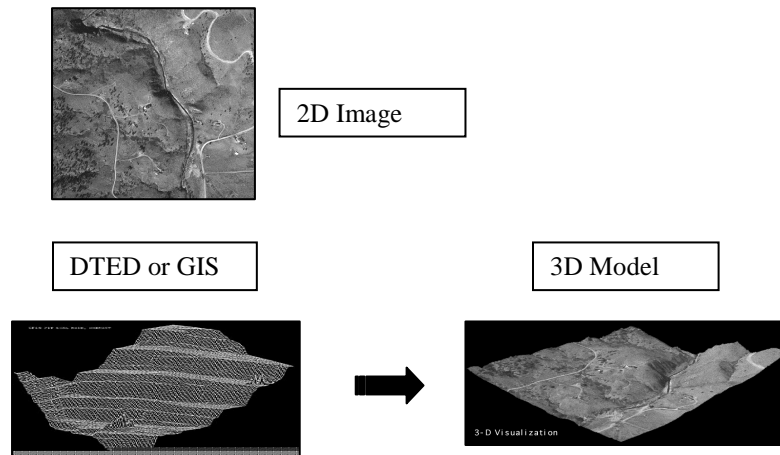
Images provided by Blasch [18]

Figure 30: 3D Model Creation. 3D model created from the combination of 2D images and geographic reference information.

Chapter Summary

Several methods of calibration and registration are available for use in stereo image processing. The method used in this calibration and the associated results of the basic calibration parameters are shown. Image registration is explained in theory, but is left up to the user to manipulate and register the images from the data sets collected. The choice of registration algorithms are dependent on the user requirements and the accomplishment of registration is beyond the scope of this thesis.

V. Conclusions and Recommendations

Conclusions

The investigation and analytical evidence provided in this thesis show the significance for gathering rich data sets which can be used in the verification of the accuracy of imaging data received from an airborne surveillance platform. One of the discrepancies in verifying the accuracy of information from airborne imaging platforms is the lack of ground-truth data. The difficulties arise due to the cost of surveying and controlling a vast area, as well as the presence of an inevitable source of error in the GPS or INS data. It is also difficult to find a large number of easily detectable landmarks used to self localize each camera. The inaccuracies are further compounded by the dynamic changes in camera orientation with respect to the aircraft frame, as it flies above the areas of interest in a circular pattern. The small-scale imaging platform could be used to study these complex issues. The small-scale platform simulates an airborne surveillance platform by capturing images in a 360 degree circle from above a known created or mock scene. The platform is not rigidly fixed and can replicate some of the flight variations, namely pitch and yaw, that an airborne platform may experience during a surveillance sortie while capturing images. Most importantly, the mock scene contains objects of known dimensions and orientation which can be used as the ground-truth data for verification of imaging algorithms. Acquisition of this kind of ground-truth verification data is hard to obtain with current airborne imaging systems in areas where the objects being viewed are unknown or where there isn't any DTED or GIS information.

The outlined research objectives of this investigation were successfully accomplished. The entire project was completed for under $250.00 and meets all of the research objectives outlined in the introduction.

A) (Objective) Modular – Hardware and software components of the system should be easily obtainable and allow for swift reconfiguration during operation.

(Objective met) – All components consist of common items found in any retail or hardware store and can be reconfigured with a variety of options due to the implementation of the IEEE 1394 (fire wire) and remote laptop computer interface.

B) (Objective) Scalable – System operating parameters and configuration should be employable at various facilities without any major modifications.

(Objective met) – The entire system can be quickly disassembled (camera rod is the only item which needs to be removed for ease of transport) and moved to various facilities which offer any type of rigging for a hanging structure – to include hoists, hard hanging points or hooks as long as the baseline camera rod has the clearance for rotation.

C) (Objective) Integration – Should abide by current FCC rules and regulations. Common electrical and computer outlets should be utilized.

(Objective met) – No FCC violations are present and all associated components operate from common electrical and computer outlets.

D) (Objective) Low Cost – Should use commercial off-the-shelf (COTS) materials.

   (Objective met) – Project designed for under $250.00 and all components are standard off-the-shelf (COTS) hardware.

E) (Objective) Low Maintenance – Design should allow for infrequent, quick repairs.

   (Objective met) – Once the setup was complete only minor adjustments needed to be made. No repairs were required during data collection.

Recommendations

While the small-scale imaging platform proved it could obtain a robust set of images from a simulated airborne platform, several modifications and fine-tuning could be made to enhance the value of the system for future work.

First, the conditions under which the platform operates could be modified. The images were obtained at a particular winch-limited-height of 6.5 feet; however, other facilities may offer different hanging fixtures which might facilitate greater platform heights. Increasing the height of the imaging platform will increase the field of view for the stereo cameras and allow for a larger scene to be created on the ground. A larger scene on the ground will allow for more objects to be placed in the scene and an increase in the data points to be collected for analysis.

Various lighting conditions could also be explored. An investigation into how lighting affects object recognition and the accuracy of object position data could be accomplished. In addition, experimenting with different objects and their placement in a scene may lead to finding weak spots in the image registration algorithms for further

study. For instance, having one object partially block out another object when the stereo pair are at a particular location and seeing if both objects could be detected and their positions found.

Lastly, the digital projector could be used as a stipe-gird projector to enhance the object detection and registration of the viewed scene. A simple Microsoft PowerPoint slide with a grid-like transmission of lines onto the scene below could be used to study and analyze its influence on the accuracy of the position data received from the stereo pair.

Bibliography

1. McNutt, R., "Project Angel Fire – GRD and OPTech '05 Development Effort," Microsoft PowerPoint presentation, Air Force Institute of Technology, Wright-Patterson AFB, Dayton, OH, 14 Oct. 2005.

2. Burt, D.J., "Basic Operation of the Charge Coupled Device," International Conference on Technology and Applications of Charge Coupled Devices, Edinburgh: University of Edinburgh, Centre for Industrial Consultancy and Liaison, 1974.

3. Fraser, C.S., Baltsavias, E., Gruen, A., "Processing of *Ikonos* Imagery for Submetre 3D positioning and Building Extraction", ISPRS Journal of Photogrammetry & Remote Sensing, v56, p177-194, 2002.

4. Boyle, W.S., Smith, G.E., "Charged Coupled Semiconductor Devices," *Bell Sys. Tech. J.*, 49:587-593, 1970.

5. Grometstein, A. A. *The Roots of Things: Topics in Quantum Mechanics*. New York: Kluwer, pp. 183-195, 1999.

6. Melen, R., Buss, D., *Charge-Coupled Devices: Technology and Applications*, New York, NY, IEEE Press, 1977.

7. Chawla, A., Pollard, B., Samei, E., Hashimoto, N., "Effect of Increased Ambient Lighting on Detectability – A Psychophysical Study," SPIE Vol 6516, 651617, Medical Imaging 2007: PACS and Imaging Informatics, 2007.

8. Hata, S., Shima, D., Kaida, K., "3-D Vision Sensor with Multiple CCD Cameras," Kagawa University, Matsushita Kotobuki Electronics Co, 2002.

9. Ko, Jung-Hwan, Lee, Jun-Ho, Kim, Eun-Soo, "Adaptive 3D Target Tracking and Surveillance Scheme based-on Pan/Tilt-embedded Stereo Camera System," 3D Display Research Center, National Research Lab of 3D Media, Dept. of Electronic Engineering, Kwangwoon University, 2004.

10. *STH-MDCS-VAR-VAR/-C Stereo Head User's Manual*, Videre Design, 2003.

11. Konolige, K., Beymer, D., Small Vision Systems (SVS) Users Manual Software Version 4.2, SRI International, Feb. 2006.

12. Lim, H., "Range Data from Stereo Images of Edge Points," SPIE Vol 1382, Intelligent Robots and Computer Vision IX: Neural, Biological and 3D Methods, pp. 434-442, 1990.

13. Bao, Hong. *Video Camera Calibration Based on Vanishing Points.* MS Thesis, Center for Advanced Computer Studies, University of Southwestern Louisiana, Lafayette, LA, July 1994.

14. Faugeras, O., *Three Dimensional Computer Vision*, Cambridge, MA, MIT Press, 1993.

15. Trucco, E., Verri, A., *Introductory Techniques for 3D Computer Vision*, Upper Saddle River, NJ, Prentice Hall, 1998.

16. *Camera Calibration Toolbox for Matlab,* http://www.vision.caltech. edu/bouguetj/calib_doc/htmls/example.html. 17 Nov. 2006

17. Zitova, B., Flusser, J., "Image registration methods: a survey," Department of Image Processing, Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic, 2003.

18. Goshtasby, A., 2-*D and 3-D Image Registration for Medical, Remote Sensing, and Industrial Applications.* Hoboken, NJ: John Wiley & Sons, Inc, 2005.

19. Blasch, E., "Fusion and Registration Interaction," Microsoft PowerPoint Presentation, Wright State University, Dayton, OH, 03 Mar. 2007.

Vita

Captain Steven A. James was born in Snoqualmie, Washington in November 1972. After living 16 years in the Northwest, his family moved to New Mexico where he graduated from Pojoaque High School in 1991 and entered the Air Force through the Delayed Enlistment Program. Upon completion of Basic Military Training at Lackland AFB, TX, and Aircraft Armament Systems Specialist School, Lowry AFB, CO, Airmen James was assigned to the 523rd Fighter Squadron, 27th Fighter Wing, Cannon AFB, NM. During his four year tenure at Cannon AFB, Airman James became a weapons loading instructor and deployed several times, to include Operation Provide Comfort in 1993, in support of the Northern no-fly zone restrictions placed upon Iraq.

In 1995, SrA James received a "Scholarships for Outstanding Airman (SOAR)" scholarship from the Air Force to become an officer and to study mechanical engineering at the college of his choice. After attending his first year of undergraduate studies at San Francisco State University and the University of California, Berkeley, Cadet James transferred to the University of New Mexico and completed B.S. in Mechanical Engineering in December 2000. He was commissioned as a officer in the Air Force in February of 2001.

2Lt James reported for Specialized Undergraduate Pilot Training (SUPT) at Laughlin AFB, TX, after a brief stint as a ROTC student recruiter, during the summer of 2001. Lt James graduated from SUPT and received his pilot wings the following year. The Air Force selected Lt James to fly the C-21A with the 457th Airlift Squadron at Andrews AFB, MD, as his first assignment. As a C-21A pilot, Captain James flew numerous high visibility missions to include the U.S. Speaker of the House, Senators and

many high-ranking officers and foreign dignitaries. He also deployed as a pilot to the Middle East in support of Operation Iraqi Freedom and Operation Enduring Freedom.

In August of 2004, Captain James was the victim of a mugging while on leave to visit friends in Chicago, IL. The doctors diagnosed him with a Traumatic Brain Injury (TBI) and he was subsequently medically disqualified from flying for the Air Force.

In April 2005 Captain James was selected to attend the Air Force Institute of Technology Graduate School of Engineering and Management for his Masters of Science in Aeronautical Engineering. He chose to focus his studies in digital avionics systems and conventional weapons and to complete a thesis in the electrical engineering and computer science discipline. Upon graduation in June 2007, Captain James will separate from the Air Force after 10 ½ years of service to his country.

1. REPORT DATE *(DD-MM-YYYY)* 14-06-07	2. REPORT TYPE Master's Thesis	3. DATES COVERED *(From – To)* June 2005 – June 2007

4. TITLE AND SUBTITLE A SMALL-SCALE IMAGING PLATFORM FOR ALGORITHM PERFORMANCE EVALUATION	5a. CONTRACT NUMBER
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S) James, Steven A., Captain, USAF	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES(S) AND ADDRESS(S) Air Force Institute of Technology Graduate School of Engineering and Management (AFIT/EN) 2950 Hobson Way, Building 640 WPAFB OH 45433-8865	8. PERFORMING ORGANIZATION REPORT NUMBER AFIT/GAE/ENY/07-01

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) NA	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

   In recent years, world events have expedited the need for the design and application of rapidly deployable airborne surveillance systems in urban environments. Fast and effective use of the surveillance images requires accurate modeling of the terrain being surveyed. The process of accurately modeling buildings, landmarks, or other items of interest on the surface of the earth, within a short lead time, has proven to be a challenging task. One approach of high importance for countering this challenge and accurately reconstructing 3D objects is through the employment of airborne 3D image acquisition platforms. While developments in this arena have significantly risen, there remains a wide gap in the verification of accuracy between the acquired data and the actual ground-truth data. In addition, the time and cost of verifying the accuracy of the acquired data on airborne imaging platforms has also increased. This thesis investigation proposes to design and test a small-scale 3D imaging platform to aid in the verification of current image acquisition, registration and processing algorithms at a lower cost in a controlled lab environment. A rich data set of images will be acquired and the use of such data will be explored.

15. SUBJECT TERMS
2D, 3D, 3D modeling, stereo imaging, image acquisition, image calibration, image registration, airborne imaging platforms, airborne surveillance systems, imaging algorithm performance evaluation, imaging benchmark

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON Dr. Guna Seetharaman
a. REPORT	b. ABSTRACT	c. THIS PAGE			19b. TELEPHONE NUMBER *(Include area code)* (937) 255-3636, ext 4612 guna.seetharaman@afit.edu)
U	U	U	UU	75	